Universidade do Minho
Escola de Engenharia

Mário João Guedes Pinto

Aquisição Rápida 360° de Informação 3D usando uma Configuração Estática

Outubro de 2014

Aquisição Rápida 360° de Informação 3D usando uma Configuração Estática

Mário João Guedes Pinto
(mariojgpinto@gmail.com)

Dissertação submetida à Universidade do Minho para obtenção do grau de Mestre em Informática, elaborada sob a orientação de

Luís Paulo Santos

Departamento de Informática
Escola de Engenharia
Universidade do Minho

Braga, Outubro, 2014

Resumo

Nesta dissertação pretende-se estudar e desenvolver um sistema de baixo custo que, através de uma configuração estática, seja capaz de realizar a captura 360° de informação 3D de objetos de pequena/média dimensão em tempo real. Este tipo de aquisição distingue-se por conseguir capturar a informação geométrica da superfície de um objeto ao longo de todo o seu perímetro exterior, isto é, conforme vista de qualquer ponto de um círculo que contenha o objeto no seu interior.

Já existem sistemas capazes de realizar este tipo de captura, no entanto, nenhum cumpre todos os requisitos acima descritos. Seja pela restrião do tamanho dos objetos suportados ou pela incapacidade de os capturar com taxas de atualização interativas, não foi encontrado nenhum sistema de baixo custo capaz de realizar uma captura 360°.

Para atingir os objetivos utilizou-se uma *Kinect* para realizar a captura e uma configuração de espelhos para fornecer ao sistema informação que o sensor não consegue capturar diretamente. Desta forma, o sistema construído é estático, nem a configuração nem o objeto necessitam de se mover para realizar a captura 360°, e consegue atingir taxas de atualização interativas com um custo reduzido.

Os resultados obtidos não tiveram o nível de detalhe pretendido, impossibilitando assim a utilização do sistema para a reprodução fiel dos objetos. A informação gerada pelo sistema respeita as formas e dimensões dos objetos, no entanto os resultados apresentam ruído e imprecisão nas superfícies dos mesmos. Quanto maior a distância a que o objeto se encontra da *Kinect*, maior a imprecisão. Como tal, as superfícies capturadas através dos espelhos obtiveram piores resultados, uma vez que a informação tem que percorrer uma distância maior para chegar ao sensor. Ainda assim, uma vez que a forma geral dos objetos se mantém e dada a velocidade de captura atingida, este sistema consegue gerar aproximações aos modelos em tempo real, permitindo ter um *input* de informação 360° para aplicações interativas a um baixo custo.

Abstract

This dissertation aims to study and develop a low cost system capable of making 3D 360° data acquisition of small/medium objects, in real time, through a static configuration. This sort of acquisition can be distinguished by capturing the geometric information of an object's surface along its entire outer perimeter, i.e., as seen from any point of a circle containing the object.

There are already some systems capable of performing this sort of capture however, none meets all the requirements aforementioned. Either by size restriction of the supported objects or the inability to capture at interactive update rates, no low-cost system capable of performing a 360° capture was found.

To achieve the proposed objectives, a *Kinect* was used to capture the information and a configuration of mirrors was built to provide the system with information that the sensor cannot capture directly. By doing so, the built system can achieve interactive update rates at a reduced cost using a static configuration: neither the mirrors nor the object need to be moved to perform the 360° capture.

The obtained results did not achieve the desired level of detail making it impossible to use the system for a reliable reproduction of objects. The information generated by the system respects both shapes and dimensions of the objects. However, the results show noise and inaccuracy on their surfaces. The greater the distance of the object from the *Kinect*, the greater the inaccuracy of the results. Thus, the areas captured by the mirrors had worse results since the information has to travel a greater distance to reach the sensor. Nevertheless, since the overall shape of the objects is maintained and considering the speed achieved, this system is able to generate approximations to the models in real time, allowing interactive applications to have a 360° information input at a low cost.

Agradecimentos

Gostaria de agradecer, em primeiro lugar, ao professor Luís Paulo Santos, não só pela orientação prestada ao longo de todo este percurso, mas também, por toda a disponibilidade e liberdade dadas desde o ínicio do projeto (e pela *kinect* que me permitiu fazer tantos testes).

Agradeço ainda ao *CCG* pela disponibilização de tempo, material e espaço para a realização dos testes. Agradeço também ao *Ido Iurgel* pela ajuda e debates pré-tese e por me ter mostrado novas àreas para aplicação de conhecimentos.

Agradeço ainda a todos os amigos e colegas, de trabalho e de lazer, que estiveram sempre prontos a distrair-me e que fizeram surgir novas ideias e projetos.

Por fim, agradeço aos meus pais, avós e restante família, que sempre me apoiaram e ajudaram durante todo o caminho, ainda mais nesta última fase, e à Margarida Sousa, por toda a ajuda, alegria e companhia que tornaram esta jornada mais fácil.

Conteúdo

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

Todos os dias interagimos com objetos sem precisar de prestar especial atenção ao posicionamento dos mesmos. A perceção da geometria destes objetos e a distância a que estes estão é feita de forma tão natural que nem nos apercebemos disso. No entanto, no mundo digital, fazer com que um computador reconheça autonomamente essa informação sobre as cenas não é um assunto trivial.

Quando capturado com uma câmara convencional (RGB), o mundo 3D é mapeado numa imagem 2D e a noção de profundidade (pode) perder-se com alguma facilidade, dificultando o reconhecimento da correta geometria dos objetivos.

No caso dos *scanners* 3D essa informação já existe, o que ajuda à análise da cena e dos objetos nela presente. Este tipo de tecnologia está em constante evolução e há cada vez mais sistemas capazes de o fazer e cada vez mais acessíveis: a *Kinect* é um exemplo disso. Este tipo de sensores normalmente é direcional, isto é, apenas conseguem adquirir aquilo que é visível e não conseguem tratar das partes ocludidas, como tal, não conseguem adquirir toda a geometria do objeto. Contrariamente a este tipo de captura, entende-se por aquisição 360° a captura de informação geométrica sobre a superfície de um objeto ao longo de todo o seu perímetro exterior, isto é, conforme vista de qualquer ponto de um círculo que contenha o objeto no seu interior. Neste tipo de captura, a informação tem de ser adquirida de várias perspetivas de forma a gerar dados que permitam posteriormente uma visualização livre do modelo capturado. No entanto, os sistemas que permitem fazer este tipo de capturas exigem a utilização de um maior número de câmaras (no caso de uma configuração estática), o que implica um maior custo, ou o posicionamento da câmara ou objeto em diferentes perspetivas (no caso de uma configuração móvel), o que implica um maior tempo para a realização da captura.

A captura de informação em 3D pode ter diferentes funções. Por exemplo,

no caso de uma aplicação de Realidade Aumentada em que são adicionados artefactos digitais à informação capturada do mundo real, se esta informação for colocada no sítio errado, os resultados dessa aplicação passam a ser incorretos e como tal podem provocar uma experiência pobre para o utilizador. Desta forma, a existência de informação volumétrica pode ser vantajosa dando uma maior robustez a este tipo de sistemas e minimizando erros de posicionamento ou *tracking*. Outro uso potencial da informação em 3D de objetos é a modelação dos mesmos para posterior utilização num mundo digital. Seja no âmbito de jogos, cinema de animação ou até mesmo em simulações, em áreas tão diversas como indústria, medicina ou militar, modelos de objetos são usados com diferentes fins e como tal, sistemas que possam ajudar na construção dos mesmos são desejáveis. Outra possível utilização destes modelos é a replicação através das impressoras 3D. A ponte entre estas duas áreas, aquisição e reprodução de informação, é aliciante e pode abrir portas a novas possibilidades.

1.1 Motivação

A utilização de informação em 3D é útil e pode ser usada de diferentes formas. O caso específico da captura 360° da informação de um determinado objeto, além de útil levanta um desafio interessante relativamente ao método da sua realização.

A utilização de uma configuração móvel, isto é, uma configuração em que é necessário haver o movimento da câmara ou do objeto para a recolha de toda a informação do mesmo, faz com que o processo possa ser mais barato, pois apenas é necessária uma câmara, mas lento, uma vez que a captura é feita de forma incremental. Por outro lado, uma configuração estática permite que a captura de informação das várias perspetivas seja feita em simultâneo. Além da aquisição de informação 3D de um objeto fixo ser feita de forma mais rápida, este *setup* possibilita ainda a aquisição e geração de informação em 3D das várias perspetivas em tempo real, podendo neste caso ambicionar-se a captura de 3D de entidades em movimento. No entanto, este tipo de configuração é por norma mais dispendioso uma vez que é frequente utilizar múltiplas câmaras para realizar a captura.

Desta forma e juntando o melhor das duas configurações, o desafio será conceber um sistema capaz de fazer uma aquisição 360° em tempo real, utilizando apenas uma câmara e uma configuração estática. Isto permitirá a aquisição de mais informação em menos tempo e de uma forma económica. Um sistema deste género permitirá ainda a geração de vídeo em 3D a 360° que posteriormente, no prisma do espectador, possibilitará a visualização da informação de forma dinâmica e a partir de várias perspetivas. A geração de dados com

estas características poderá também ser usada em sistemas holográficos uma vez que existe a informação necessária para uma visualização livre do objeto em foco.

1.2 Objetivos

Pretende-se com esta dissertação desenvolver um sistema de baixo custo, com apenas uma câmara, capaz de realizar a captura 360° de informação 3D em tempo real a partir de uma configuração estática. Inicialmente será necessário compreender o processo de captura 3D e como a partir dessa informação poderemos gerar a representação do objeto em questão.

A captura 360° de informação 3D de um objeto requer a aquisição de informação de várias perspetivas. Como foi referido, para o conseguir fazer de forma instantânea (todas as perspetivas ao mesmo tempo) é comum usar-se vários sensores/câmaras, no entanto, além do aumento da carga computacional, isso também envolve um maior custo em *hardware*. Desta forma, um dos desafios será conceber a arquitetura de um sistema que consiga obter a informação de todas essas perspetivas em simultâneo e de forma eficiente utilizando apenas um sensor. Dada a captura e tratamento de informação em 3D envolver um grande esforço computacional, outro desafio será estudar formas de agilizar este processamento.

Em suma, os objetivos para esta dissertação são:

- Compreender os diferentes processos de captura de informação de 3D.
- Estudar e conceber uma arquitetura do sistema capaz de recolher a informação do objeto de todas as perspetivas utilizando apenas uma câmara estática e sem mover o objeto.
- Desenvolver esse sistema de captura e também a componente de visualização de informação em tempo real.
- Estudar formas de melhorar o desempenho e a qualidade da informação capturada.

1.3 Estrutura do documento

Neste capítulo foi introduzido o tema e a ideia motora desta tese e foram apresentados os objetivos da mesma. Nos próximos capítulos será mostrado qual o caminho seguido, desde a conceção do sistema, a sua implementação,

até aos resultados e considerações finais. A estrutura deste documento é a seguinte:

- Capítulo 2 - Estado da Arte: Descreve o Estado da Arte na área da digitalização 3D, com foco na aquisição 360°. Aqui serão mostrados os principais métodos de captura 3D, sistemas de aquisição 360° já existentes e aplicações desta tecnologia.

- Capítulo 3 - Visão do Sistema: Analisa em mais detalhe quais os objetivos desta tese de forma a delinear qual o caminho a seguir. Aqui serão apresentadas as decisões tomadas relativamente ao sensor e configuração a utilizar e serão ainda expostos alguns possíveis casos de uso para o sistema a desenvolver.

- Capítulo 4 - Implementação: Relata os passos seguidos para a implementação do sistema. É definida a arquitetura do sistema e mostrado o fluxo de execução do mesmo, apresentando depois alguns dos problemas que surgiram durante o desenvolvimento, bem como as soluções implementadas. São ainda apresentadas de forma breve as tecnologias utilizadas no âmbito deste projeto.

- Capítulo 5 - Resultados: Apresenta e analisa os resultados obtidos. São expostos os métodos de avaliação, bem como os objetos em estudo e, de seguida, os resultados serão analisados e validados e para aferir a qualidade dos mesmos. O desempenho do sistema é também apresentado neste capítulo.

- Capítulo 6 - Conclusão e trabalho futuro: Finaliza esta dissertação fazendo um resumo geral do trabalho realizado com foco nos resultados obtidos. São ainda apresentadas possíveis melhorias e adições ao sistema.

Capítulo 2

Estado da Arte

Os primeiros sistemas de captura de informação em 3D remontam à década de 1960 e estes usavam luzes, câmaras e projetores para realizar a tarefa [Lerch et al., 2006]. Era um processo moroso que exigia muito esforço e tempo para conseguir ter resultados satisfatórios. Durante vários anos esta tecnologia não sofreu grandes desenvolvimentos e tal pode também ser justificado, por exemplo, pelas limitações de largura de banda ou pela capacidade de armazenamento disponível. No fim dos anos 1980 foram criados os primeiros *scanners* 3D a laser que usavam luz branca, lasers e sombras para capturar a superfície de objeto.

Desde então a tecnologia tem evoluído a passos largos e têm surgido vários sistemas utilizando técnicas diferentes para o mesmo fim: fazer a digitalização de informação 3D. Estas diferentes técnicas conferem diferentes características aos sensores permitindo que estes sejam usados com objetivos diferentes como a captura a longa ou curta distância, a aquisição de uma qualidade detalhada ou a preferência pela prototipagem rápida, etc. O aperfeiçoamento e difusão destes instrumentos serviu também como alavanca para algumas áreas como a Antropometria [Jones and Rioux, 1997] ou a preservação digital [Berndt and Carlos, 2000][Henry et al., 2010].

Atualmente existem vários dispositivos capazes de fazer aquisição 3D de forma fácil e rápida. Além dos sensores industriais orientados a capturas de grandes dimensões existem também sistemas que permitem fazer esse tipo de aquisições em ambiente doméstico. Produtos como o *Matterform*[1] e o *Digitalizer*[2] permitem capturar modelos 3D de objetos de pequenas dimensões com grande qualidade (erros nas ordens dos milímetros) e em poucos minutos. A *Kinect* é outro exemplo que prima pela sua versatilidade e tanto consegue capturar a geometria dos objetos como também realizar a captura e criar o

[1] `https://matterandform.net/` (acedido em outubro de 2014)
[2] `http://store.makerbot.com/digitizer.html` (acedido em outubro de 2014)

modelo de um cenário completo. A vertente móvel deste tipo de sensores tem-se tornado cada vez mais apelativa. Soluções como o *Capri* da *Primesense*[3] ou projetos como o *Structure Sensor*[4] ou o *CADScan*[5] poderão levar à proliferação destes sensores em ambiente de mobilidade o que por sua vez impulsionará avanços no campo da Realidade Aumentada.

Neste capítulo serão descritos diferentes métodos de captura de informação 3D, focando a análise nos processos óticos de Estereoscopia, *Time-of-Flight* e Luz Estruturada uma vez que são os mais comuns e os que mais se enquadram com a ideia proposta. Será também abordada a *Kinect* enquanto marco da captura de informação tridimensional. Depois disso serão mostrados alguns exemplos de aplicações que conseguem realizar a captura 360° e por fim serão listadas algumas das possíveis aplicações práticas dos *scanners* 3D.

2.1 Métodos de captura

Existem várias tecnologias que podem ser usadas para fazer a captura de informação 3D e não existe uma que seja melhor que as outras. Todas elas têm vantagens e desvantagens de acordo com os objetivos pretendidos, além de um custo associado, que varia consoante as características dos sistemas. Se o objetivo for reconstrução de artefactos arqueológicos, a captura não poderá ser invasiva e o nível de detalhe terá que ser elevado, no entanto o tempo de aquisição não será uma limitação. Por outro lado, se se tiver um sistema de videoconferência ou de interação em tempo real, o tempo de aquisição terá que ser mínimo, enquanto que a qualidade dos modelos passa a ser secundária e pode ser até aproximada a modelos já conhecidos.

Os diferentes requisitos das aplicações fizeram com que fossem criados diferentes tipos de sistemas de captura. O esquema presente na Figura 2.1 representa uma taxonomia dos sensores de aquisição 3D.

[3] http://www.i3du.gr/pdf/primesense.pdf/ (acedido em outubro de 2014)
[4] http://structure.io/ (acedido em outubro de 2014)
[5] http://cad-scan.co.uk/ (acedido em outubro de 2014)

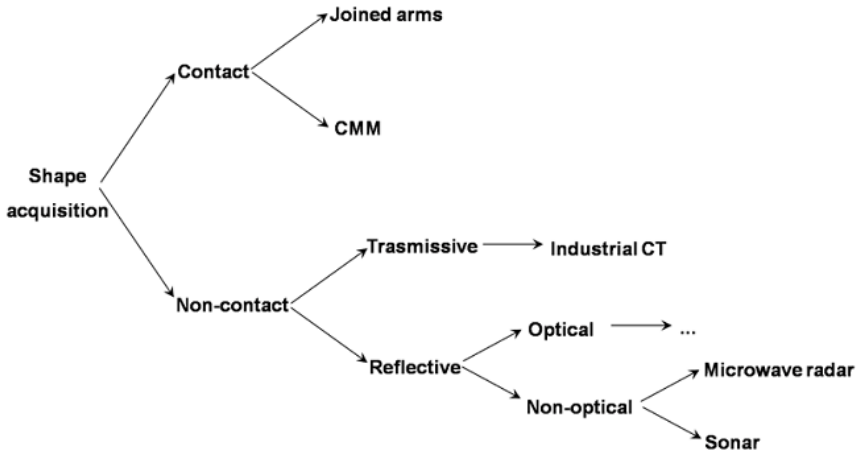


Figura 2.1: Taxonomia dos sensores de aquisição 3D[6].

A primeira grande divisão prende-se com o fato de a captura ser feita, ou não, através de contacto. Por norma, técnicas de captura 3D com contacto são intrusivas e, como tal, podem alterar as características físicas do objeto. As técnicas mais usadas neste campo são a *CMM* (*Coordinate Measuring Machine*) e o *Jointed Arm*. Do outro lado, as técnicas que não utilizam o contacto, funcionam a partir da interação entre a superfície do objeto e algum tipo de radiação. Contudo, no caso da radiação do tipo transmissiva, esta também é intrusiva, uma vez que é absorvida pelo objeto e, como tal, pode alterar as suas propriedades. Exemplos deste tipo de sistemas são a Tomografia Computadorizada (*CT*) e a Ressonância Magnética.

As tecnologias reflexivas, tal como o nome indica, exploram a radiação refletida pelos objetos para inferir a posição dos pontos. Estes sistemas não são intrusivos e, dependendo do comprimento de onda utilizado pela radiação, podem ser divididos em sistemas óticos ou não-óticos. Os sistemas não-óticos utilizam radiação eletromagnética, ondas sonoras ou ultra-sons para realizar a captura, e baseiam-se no princípio de *time-of-flight* para efetuar as medições.

O caso das tecnologias óticas requerem um maior detalhe uma vez que são as mais comuns, e as que mais se adaptam ao sistema proposto para esta dissertação (Figura 2.2). Usam radiação no espetro visível (100nm - 300um) e são, por norma, tecnologias de baixo-custo que permitem realizar a captura de informação com boa qualidade, de forma rápida, e em grande escala. No entanto, um problema deste tipo de sistemas é, por exemplo, a dificuldade em capturar a informação de superfícies brilhantes, reflexivas ou transparentes ou ainda, as oclusões causadas pela própria geometria do objeto ou por diferentes objetos.

[6]Retirado de [Bellocchio and Ferrari, 2011].

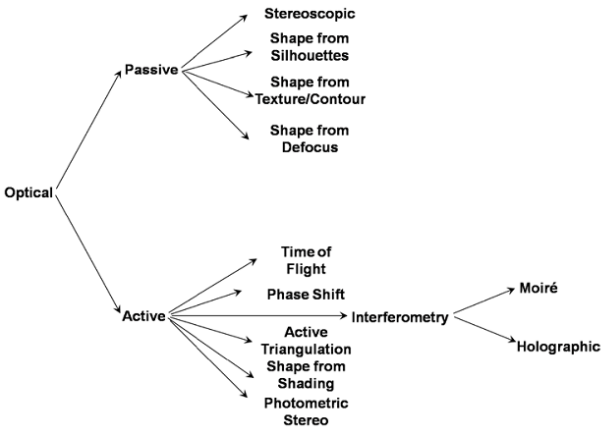


Figura 2.2: Taxonomia dos sensores de aquisição 3D óticos[7].

Dentro das tecnologias óticas, estas podem ser divididas em passivas ou ativas. Os sistemas passivos não emitem qualquer tipo de radiação e usam apenas aquela que é refletida pelo ambiente. Estas técnicas utilizam configurações óticas e/ou perspetivas diferentes que depois são processadas para gerar as coordenadas dos pontos. Estes são por norma sistemas baratos mas, em contrapartida, algo limitados, e não produzem resultados com grande detalhe. Por outro lado, nos sistemas óticos ativos já existe a emissão de radiação. A informação 3D é obtida a partir do processamento das características da radiação emitida e da informação capturada, resultado da interação entre essa radiação e a superfície do objeto. Estes sistemas são os mais comuns sendo que, uma das razões é a facilidade em capturar não só a informação 3D mas também as cores dos objetos em simultâneo.

As divisões apresentadas até agora representam caraterísticas das tecnologias existentes, no entanto há sistemas que usam mais que uma tecnologia. Estes sistemas híbridos são por norma mais robustos e apresentam resultados mais precisos à custa de uma maior complexidade e de um preço mais elevado.

De seguida serão abordados em mais detalhe os sistemas de captura de Estereoscopia, *Time-of-Flight* e Luz Estruturada uma vez que são os sistemas mais usados atualmente [Ko and Agarwal, 2012] e os que mais se adaptam ao desafio presente.

2.1.1 Estereoscopia

De acordo com a taxonomia apresentada anteriormente, a estereoscopia é um método de captura de informação 3D que não utiliza contacto, é refletivo, ótico e passivo. Esta técnica baseia-se no sistema visual humano, isto é, um

[7] Retirado de [Bellocchio and Ferrari, 2011].

sistema bi-ocular. Os olhos humanos estão separados por uma distância de aproximadamente 6,5cm, o que faz com que tenham perspetivas ligeiramente diferentes da mesma cena. Estas duas imagens são processadas pelo cérebro e é a diferença entre elas que permite ter a noção de profundidade e, desta forma, saber assim a que distância é que um determinado objeto se encontra (Figura 2.3).

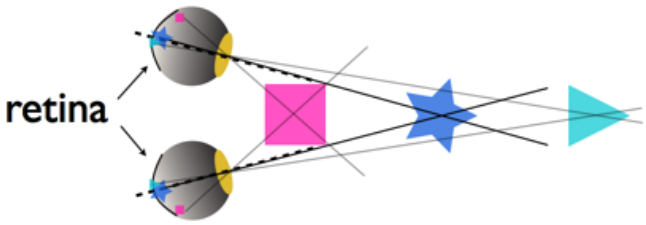


Figura 2.3: Exemplo de visão estereoscópica e da noção de profundidade, de acordo com o sistema visual humano[8].

Os sistemas de captura estereoscópica baseiam-se no mesmo princípio para calcular o mapa de profundidades. São utilizadas imagens de uma determinada cena, mas com perspetivas diferentes, e a partir delas é construído um mapa de disparidades, isto é, um mapa com a representação das diferenças entre as duas imagens. Tendo essa informação, juntamente com a informação relativa às camaras, é possível estimar a posição 3D dos pontos da cena através por triangulação. Numa fase prévia à captura, são obtidas características como a posição relativa da câmara, orientação e os parâmetros internos (distância focal, centro ótico, parâmetros de distorção, etc.). A triangulação é feita para todos os pontos que têm correspondência nas duas imagens e, para cada um (dos pontos), é projetado um raio de acordo com as características recolhidas na fase de calibração. A interseção destes dois raios contém a representação 3D do ponto em questão (Figura 2.4).

O principal desafio da estereoscopia é o sistema de correspondência entre os pontos das duas imagens. Por norma, este método é apenas utilizado para a reconstrução de certos objetos com características fortes como cantos ou arestas bem definidas e em que as correspondências possam ser facilmente reconhecidas. Esta técnica requere um esforço computacional grande e a qualidade da captura está intimamente ligada à qualidade da fase de configuração/calibração e à qualidade dos sensores.

Atualmente existem vários instrumentos capazes de fazer captura estereoscópica. Estes sistemas podem ter várias aplicações pelo que, as que mais se

[8]Adaptado de
http://www.dashwood3d.com/blog/beginners-guide-to-shooting-stereoscopic-3d/.
[9]Retirado de [Bellocchio and Ferrari, 2011]

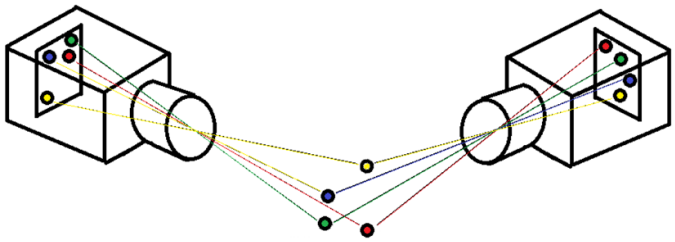


Figura 2.4: Representação da obtenção da informação tridimensional num sistema estereoscópico[9].

distinguem são, a realização de filmes em 3D, a inclusão em sistemas de robótica, e a construção de sistemas de entretenimento 3D *low-cost* como consolas e telemóveis (Figura 2.5). O facto de serem uma tecnologia com um custo reduzido fez com que a estereoscopia não estagnasse na área industrial e com que a passagem para o mercado acontecesse com alguma facilidade. De salientar a inclusão destes sistemas em consolas de jogos portáteis como a *Nintendo 3DS* que, além de possibilitar a captura de fotografias em 3D, também tira partida da visão estereoscópica para a criação de melhores sistemas de realidade aumentada.

Figura 2.5: Exemplos de produtos comerciais capazes de realizar captura estereoscópica. À esquerda, a Nintendo 3DS*, ao centro, câmara de vídeo estereoscópica profissional e à direita, câmara estereoscópica comercial*[10].

2.1.2 *Time-of-Flight*

O princípio por trás desta tecnologia baseia-se na medição da distância aos pontos das superfícies através do tempo que a radiação emitida demora a chegar

[10]Várias imagens. Da esquerda para a direira, retiradas de: http://reviews.cnet.co.uk/portable-gaming/nintendo-3ds-review-50000079/ (acedido em outubro de 2014), http://www.olivieris.toile-libre.org/index.php?pg=18&id=18 (acedido em outubro de 2014), http://tctechcrunch2011.files.wordpress.com/2009/07/3d_camera_07201.jpg (acedido em outubro de 2014).

aos objetos e voltar. Sabendo esse tempo, a velocidade, e a direção da emissão da radiação, é possível saber a distância a que uma determinada superfície se encontra e quais as coordenadas 3D dos pontos. Os sistemas *Time-of-Flight* (*ToF*) não utilizam contacto, são reflexivos e, dependendo da radiação usada, podem ser não-óticos, como os radares (ondas eletromagnéticas ou de baixa frequência), ou os sonares (ondas acústicas), ou podem ser classificados como óticos ativos no caso dos radares óticos. Estes sistemas podem ainda ser referidos como *LIDAR* (*LIght Detection and Ranging*) ou *LADAR* (*Laser Detection and Ranging*) [Bellocchio and Ferrari, 2011].

No caso dos sensores *ToF* ponto a ponto, a distância de um determinado ponto na cena é calculado pelo princípio explicado em cima ou seja, a distância à câmara de cada ponto ρ é dada por:

$$\rho = (C \times T)/2$$

Onde C é a velocidade da radiação e T o tempo medido correspondente ao percurso da radiação (Figura 2.6).

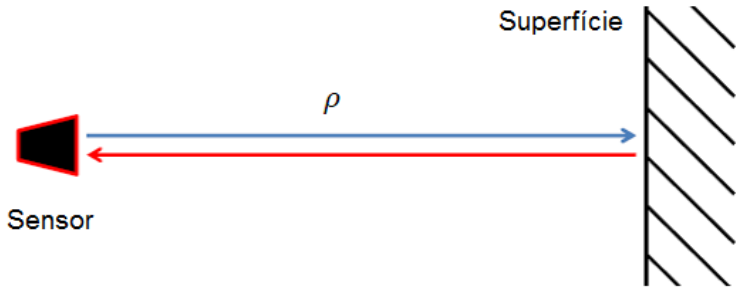


Figura 2.6: Representação do sistema de captura Time-of-Flight.

Este processo é efetuado para todos os pontos que se pretende medir, o que, normalmente, implica a movimentação do sensor, tornando-o assim inapto para capturas dinâmicas, isto é, de cenas em que existe movimento. Por outro lado, no caso dos sensores *ToF* matriciais, a geometria é capturada num único instante por uma matriz de sensores. Cada elemento dessa matriz faz a medição de forma independente produzindo um mapa de profundidade a velocidades interativas [Mutto et al., 2012].

Uma vez que este tipo de tecnologia envolve a velocidade da luz ($3 \times 10^8 m/s$), o detalhe deste tipo de tecnologias fica limitada à velocidade a que os sensores conseguem fazer essas medições. Por exemplo, para se medir um detalhe com 1mm, a diferença entre medições é de $5ps$, o que exige a existência de um relógio capaz de medir passos dessa grandeza. A escolha de diferentes tipos de relógios leva a diferentes tipos de sensores *ToF*, pelo que os mais comuns

são a abordagem de modulação da intensidade de ondas contínuas, obturadores óticos e *Single photon avalanche diodes*[Mutto et al., 2012]. Outros fatores, como a iluminação externa do ambiente ou interferências de outras câmaras, também causam perturbações na captura levando a medições imprecisas ou erradas.

Este tipo de sensores tem uma utilização bastante versátil uma vez que a amplitude do seu alcance é muito grande. Sistemas como o *FARO 3D imager* ou o *Riegl VZ-6000* (Figura 2.7) foram desenvolvidos para longo alcance, podendo efetuar medições a várias centenas de metros. Estes sistemas são ideais para fazer a medição de terrenos, tanto no chão como a partir de meios aéreos, de forma a produzir mapas topográficos. A arquitetura e construção civil também são áreas que podem beneficiar desta tecnologia através de medições e consequente validação em diferentes fases de projetos. No caso da área de preservação cultural e arqueologia estes sistemas são também utlizados para construir modelos 3D com grande detalhe e de forma não intrusiva.

Figura 2.7: Exemplos de sensores Time-of-Flight *industriais. À esquerda, a* RIEGL VZ-6000 *e à direita a* FARO 3D imager[11].

Por outro lado, sensores como a *CamCube* , a *SR4000* ou a *Intel RealSense* (Figura 2.8) foram desenvolvidos para curto alcance $(0.8 - 5m)$, atingindo velocidades de captura superiores aos $30fps$. Estas características fazem com que os sensores consigam detetar e seguir gestos efetuados por humanos com facilidade, permitindo a criação de interfaces naturais, úteis para áreas como o entretenimento, robótica ou até medicina. A possibilidade de aquisição de modelos também pode ser explorada, podendo tirar-se partido da aquisição em

[11]Várias imagens. Da esquerda para a direira, retiradas de: http://www.riegl.com/nc/products/terrestrial-scanning/produktdetail/product/scanner/33/ (acedido em outubro de 2014),
http://www.faro.com/en-us/products/metrology/faro-3d-imager/overview (acedido em outubro de 2014).

tempo real para beneficiar a captura de ambientes voláteis ou em movimento.

Figura 2.8: Exemplos de sensores Time-of-Flight *comerciais. À esquerda, a* CamCube *da* PMD Technologies*, ao centro a* SR4000 *da* Mesa Imaginig *e à direita a* Intel RealSense *da* Intel[12].

2.1.3 Luz estruturada

Os sistemas de luz estruturada enquadram-se na categoria de métodos de captura que não utilizam contacto, são refletivos e ativos. Este tipo de sistemas, normalmente, é composto por uma componente emissora e um ou mais sensores de captura. Para realizar este processo é utilizado um princípio similar àquele que é usado nos métodos estereoscópicos, a triangulação ativa, onde, neste caso, a perspetiva da segunda câmara é substituída pela entidade emissora. Este emissor projeta um padrão na cena que depois é capturado pelo sensor. Essa informação, juntamente com os dados recolhidos no processo de calibração, é analisada, e a partir das deslocações encontradas no padrão, são calculados os pontos 3D do mundo (Figura 2.9).

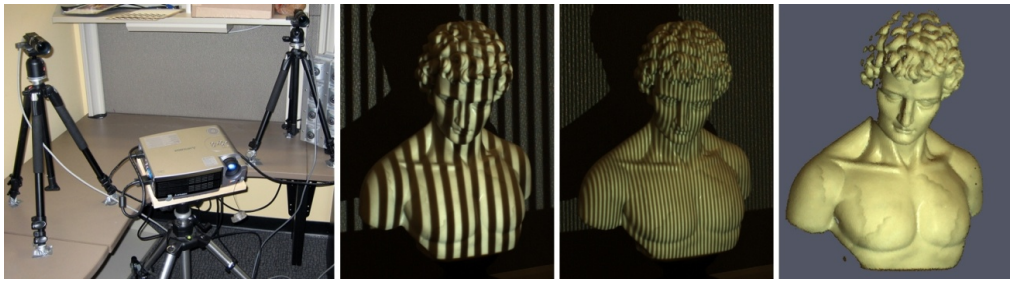


Figura 2.9: Exemplo de uma captura 3D recorrendo à técnica de Luz Estruturada[13].

[12]Várias imagens. Da esquerda para a direira, retiradas de:
http://www.pmdtec.com/ (acedido em outubro de 2014),
http://www.mesa-imaging.ch/ (acedido em outubro de 2014),
http://click.intel.com/intelsdk/Default.aspx (acedido em outubro de 2014).

Existem várias técnicas de luz estruturada que diferem essencialmente no tipo de padrão projetado o que, consequentemente, influencia a qualidade e o tempo da captura. Uma das técnicas usadas é o *Gray Code*, uma técnica sequencial de padrões binários compostos por tiras pretas e brancas seguindo a sequência de *Gray* (Figura 2.10).

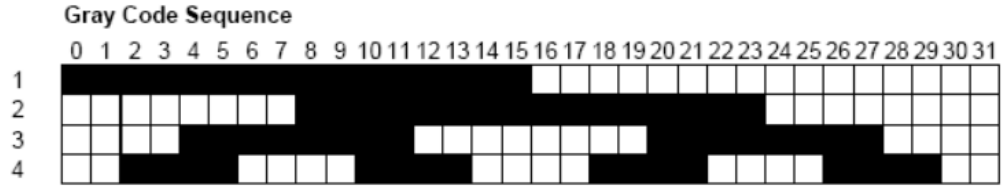


Figura 2.10: Sequência de Gray[14].

A análise dos diferentes padrões permite a identificação de zonas únicas com tiras únicas que, baseando-se no princípio de triangulação, conseguem gerar pontos 3D da cena. Esta técnica é muito precisa e bastante tolerante às texturas das superfícies uma vez que apenas usa valores binários. Em contra partida, para se conseguir alcançar resoluções elevadas, é necessário um grande número de padrões sequenciais e, como tal, a aquisição torna-se mais demorada. Isto faz com que este método seja inapto para capturas de cenas dinâmicas ou entidades vivas como o ser humano.

Outros métodos como a variação de padrões coloridos [Geng, 1996], métodos *Stripe Index* [Boyer and Kak, 1987] ou padrões de grelhas 2D espaciais [Payeur and Desjardins, 2009], permitem a aquisição da informação 3D do objeto em apenas um momento (*frame*), conferindo-lhes as propriedades necessárias para suprir os problemas descritos dos códigos sequenciais. Na Figura 2.11 são indicadas algumas dessas técnicas, de acordo com as metodologias utilizadas [Geng, 2011].

Apesar de grande variedade de métodos, e das diferentes vantagens de cada um deles, os sistemas de luz estruturada têm algumas desvantagens que são comuns entre todos. A área capturada é limitada pelo alcance do projetor e pela amplitude do sensor de captura o que, normalmente, traduz-se em áreas de ação pequenas e, como tal, o processo de captura de objetos fica limitado a objetos de pequena dimensão. Outro problema é a aquisição das características como a cor do objeto ou a luz ambiente, uma vez que a sobreposição da luz projetada adultera essas propriedades. Neste caso, o problema pode ser ultrapassado usando padrões de luz que não pertençam ao espetro visível, tal como é feito com a *Kinect* e que será explicado no Capítulo 2.1.4.

[13] Retirada de `http://mesh.brown.edu/3dpgp-2009/homework/hw2/hw2.html` (acedido em outubro de 2014).
[14] Retirada de [Geng, 2011]
[15] Retirada de [Geng, 2011]

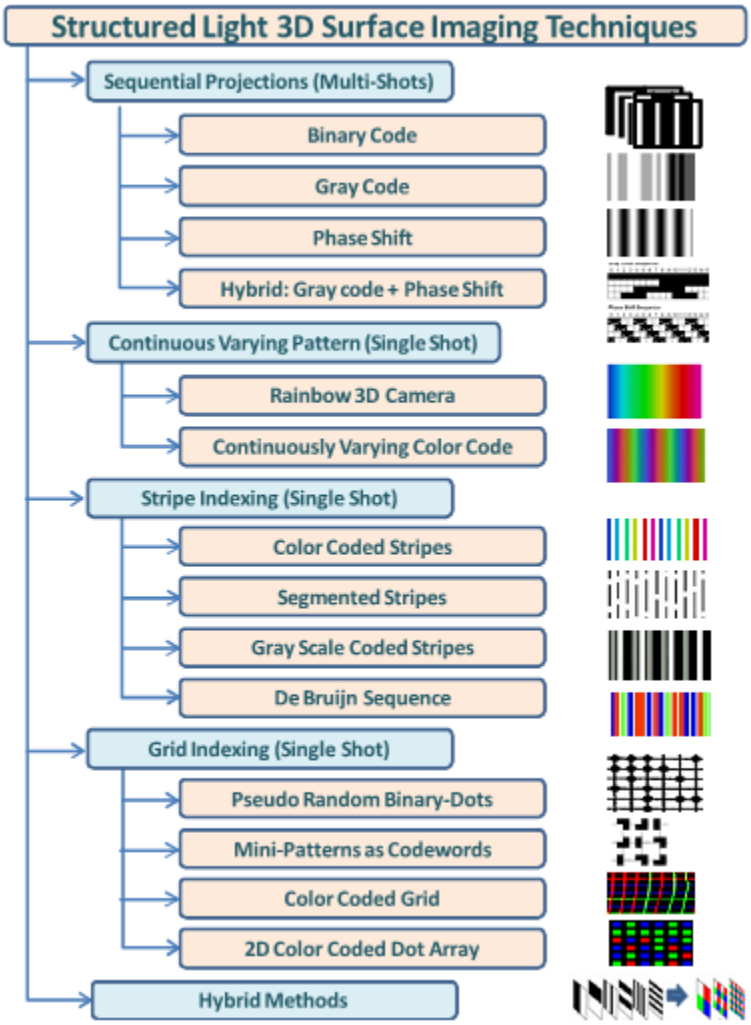


Figura 2.11: Técnicas de Luz Estruturada[15].

Estes sistemas são utilizados para a aquisição de modelos com grande detalhe nas mais diversas áreas. Várias empresas, como a *4DDynamics*[16], a *3D-Shape*[17] ou a *Createform*[18] oferecem produtos e serviços orientados para o *scan* de alta resolução do corpo-humano ou apenas da face. Isto pode ter aplicações não só em áreas como o cinema e o entretenimento mas também na saúde. Ainda nessa área, a *Maestro3D*[19] e a *Cynoprod*[20] especializaram-se

[16] http://www.4ddynamics.com/ (acedido em outubro de 2014)
[17] http://www.3d-shape.com/home/home_d.php (acedido em outubro de 2014)
[18] http://www.creaform3d.com/pt/solucoes-para-area-de-saude (acedido em outubro de 2014)
[19] http://www.maestro3d.com/ (acedido em outubro de 2014)
[20] http://www.cynoprod.com/ (acedido em outubro de 2014)

na aquisição de modelos dentários, conseguindo produzir modelos com grande detalhe de forma rápida. A *FlashScan3D*[21], mais ligada às áreas das ciências forenses, construiu um *scanner* que se destaca pela sua precisão para a captura de impressões digitais. Noutras áreas, como a indústria, estes sistemas também são usados por empresas especializadas como a *Vitronic*[22] ou a *Optimet*[23] que usam os scanners 3D para analisar peças complexas e inspecionar a sua qualidade. De salientar ainda a *Kinect* como um produto que chegou aos consumidores em grande escala, e que tem vindo a crescer o número e a variedade de aplicações que a usam, como será descrito na secção seguinte.

2.1.4 *Microsoft Kinect*

A *Kinect*, anteriormente conhecido como *Project Natal* (Figura 2.12), é uma câmara *RGBD*, ou seja, uma câmara capaz de capturar simultaneamente imagem de cor (*RGB*) e de profundidade (*(D)epth*). Foi a primeira câmara a sair para o mercado que juntou num só dispositivo esses sensores, um microfone *multi-array*, e um processador interno com *software* proprietário.

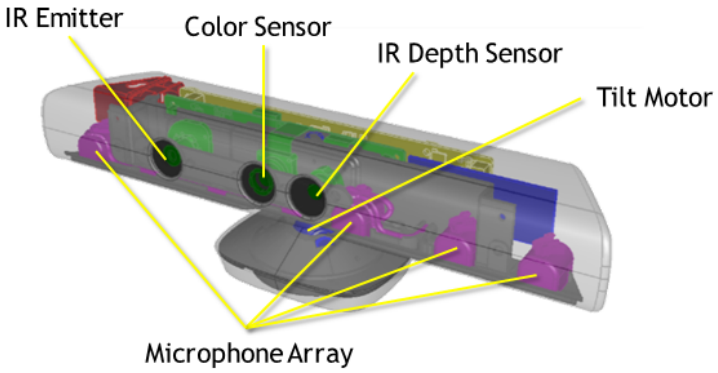


Figura 2.12: Estrutura interna da Microsoft Kinect[24].

A nível de visão, a *Kinect* vem equipada com um sensor de cor que tem a capacidade de capturar imagens *RGB* 640×480 a $30fps$ ou com uma maior resolução, 1280×960, a $10fps$. O sistema de captura de profundidade é composto por emissor e um sensor de infravermelhos. Apesar dos resultados serem semelhantes aos de sistemas de *Time-of-Flight*, a *Kinect* recolhe a informação de profundidade através da técnica de luz estruturada no espetro infravermelho. É emitido um padrão de pontos esparsos pelo emissor *IR* que depois de

[21] http://www.flashscan3d.com/ (acedido em outubro de 2014)
[22] http://www.vitronic.de/en (acedido em outubro de 2014)
[23] http://www.optimet.com/ (acedido em outubro de 2014)
[24] Retirada de http://msdn.microsoft.com/en-us/library/jj131033.aspx

analisado produz o mapa de profundidade, com uma dimensão de 320×240 e um nível de detalhe de $11 bits$ (Figura 2.13) [Mutto et al., 2012]. Informação mais detalhada sobre o funcionamento e limitações da *Kinect* será abordada na Secção 3.3.2.

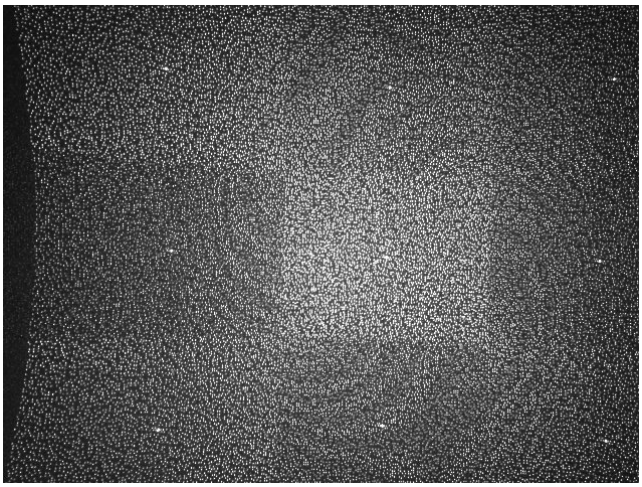


Figura 2.13: Padrão de Luz Estruturada utilizado pela Kinect[25].

Apesar de não ser o único sensor *RGBD* com este tipo de características presente no mercado (existe também a *Asus Xtion Pro Live*[26] e a *PrimeSence Carmine 1.09*[27], a *Kinect* foi a que teve uma maior aceitação. Este sensor foi desenvolvido e lançado pela *Microsoft* no final de 2010 enquanto sensor de movimento e como acessório da *Xbox 360*. Foi publicitado como uma nova forma de interação no mundo dos jogos pois não haveria a necessidade da utilização de um controlador físico, apenas faria uso de gestos e voz. Este dispositivo é um fenómeno de popularidade tendo atingido a marca de 1 milhão de unidades vendidas em menos de duas semanas (apenas nos Estados Unidos) e a marca de 8 milhões nos primeiros 60 dias, o que lhe valeu a distinção de *Fast Selling Gaming Peripheral* no livro de recordes do *Guinness*[28].

No entanto, a popularidade deste dispositivo não se ficou pelos jogos da consola e também suscitou grande curiosidade no mercado de desenvolvimento de *software* devido às suas características. A junção da informação entre câmara de cor e a de profundidade possibilitou a deteção de utilizadores na cena e inferir os seus esqueletos, com 20 pontos de controlo, em tempo real. A sua eficácia, precisão, e baixo custo, tornaram-na num objeto de eleição para

[25] Retirada de [Cruz et al., 2012]

[26] http://www.asus.com/Multimedia/Xtion_PRO_LIVE/#specifications (acedido em outubro de 2014)

[27] http://www.i3du.gr/pdf/primesense.pdf (acedido em outubro de 2014)

[28] http://www.guinnessworldrecords.com/records-9000/fastest-selling-gaming-peripheral/ (acedido em outubro de 2014)

projetos em áreas tão diferentes como robótica, jogos, interfaces naturais ou *performances* artísticas.

Quando saiu, a 4 de novembro de 2010, a *Kinect* foi lançada enquanto acessório de uma consola de jogos e não foi disponibilizado nenhum software para desenvolvimento. Nessa mesma data e de forma a contornar essa situação, a empresa *Adafruit Industries* ofereceu uma recompensa[29] para quem desenvolvesse um driver opensource capaz de aceder à informação do sensor. Seis dias mais tarde e depois de a recompensa ter triplicado, foi lançada a primeira versão do *libfreenect*, um *driver* livre capaz de ler o *stream* de vídeo de cor e de profundidade da câmara. Um mês mais tarde, e em resposta ao crescente interesse e número de projetos da comunidade opensource, a *PrimeSence* lançou o seu próprio *driver*, também ele *opensource*, a *framework OpenNI*, e ainda os binários para o *middleware NiTE*, capaz de detetar e fazer o *tracking* do esqueleto, entre outras funcionalidades. A *Microsoft* apenas entrou neste mercado em fevereiro do ano seguinte lançando um *SDK* para fins não comerciais.

Desde então os *softwares* evoluíram oferecendo maior precisão e mais funcionalidades (*tracking* de mãos, deteção de gestos, reconhecimento de voz, etc) e o *hardware* também acompanhou. E maio de 2012 foi lançada a *Kinect for Windows*, um sensor praticamente igual ao original e orientado ao mercado dos computadores pessoais. Este sensor tinha apenas como diferença a adição da funcionalidade de *Near Mode* que, respondendo ao seu novo propósito, lhe dá uma maior precisão a distâncias mais próximas da câmara. A próxima iteração na evolução neste *hardware* aconteceu em 2014, com o lançamento da *Kinect2*, juntamente com a nova geração da consola da *Microsoft*. Este dispositivo já tem uma imagem de cor *FullHD* a 30*fps*, um maior campo de visão (70° horizontal, 60° Vertical) e uma *stream* de profundidade com uma maior resolução (512×424). A qualidade da informação de profundidade capturada foi melhorada, notando-se principalmente nas zonas de conflito, como as arestas e os objetos que se encontrem mais próximos da câmara.

2.2 Sistemas de aquisição 360°

Como foi descrito na introdução, a aquisição 360° corresponde à captura de informação de várias perspetivas centradas no objeto a capturar. Para realizar este tipo de captura é preciso no mínimo de duas perspetivas diferentes do objeto. Como se pode ver na Figura 2.14, se o objeto tiver uma geometria simples, como é o caso de uma esfera, tendo apenas duas perspetivas consegue-se capturar a informação de toda a superfície do objeto. No entanto, como

[29] http://www.adafruit.com/blog/2010/11/04/the-open-kinect-project-the-ok\-prize-get-1000-bounty-for-kinect-for-xbox-360-open\-source-drivers/ (acedido em outubro de 2014)

está ilustrado no segundo exemplo da mesma figura, caso o objeto a capturar seja um pouco mais complexo, tal já não é possível e, independentemente da posição em que se irá capturar a informação, haverá zonas do objeto em que não será possível capturar informação.

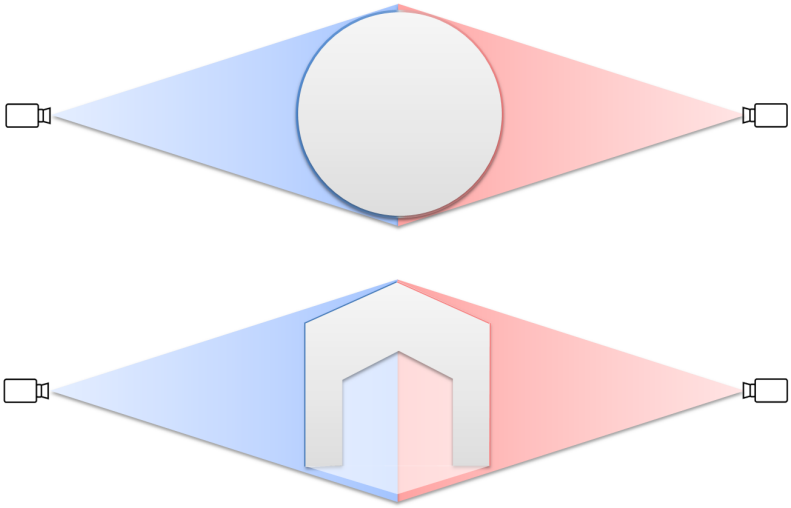


Figura 2.14: Exemplo da visualização de um objeto por duas câmaras. Na figura de cima toda a superfície do objeto é capturado enquanto que na imagem em baixo isso é impossível de acontecer.

Neste caso específico, o problema poderia ser ultrapassado com a adição de mais uma fonte de captura (Figura 2.15) e, desta forma, toda a superfície do objeto voltaria a estar coberta. Contudo, com o aumentar da complexidade do objeto ou até com a existência de múltiplos objetos em cena, esta dificuldade voltará a surgir e haverá casos em que a adição de novas perspetivas para aquisição já não conseguirão resolver a situação. Pode haver partes dos objetos que não é possíveis capturar e este problema apenas poderá ser ultrapassado com outro tipo de sistemas de captura, e não apenas os óticos.

2.2.1 Aquisição estática

Os pressupostos da aquisição estática de um objeto são que, tanto o sistema de aquisição como o objeto a capturar, não necessitam de se mover para a realização da captura 360°. Esta limitação requere a introdução de múltiplas fontes de captura, o que, por norma, implica a aquisição de mais câmaras. Além de introduzir mais complexidade e carga computacional, isso torna o custo do sistema mais elevado. Por outro lado, com uma configuração estática, tem-se como vantagem a possibilidade de aquisição de informação das diferentes perspetivas em simultâneo. Esta característica beneficia as capturas de entidades

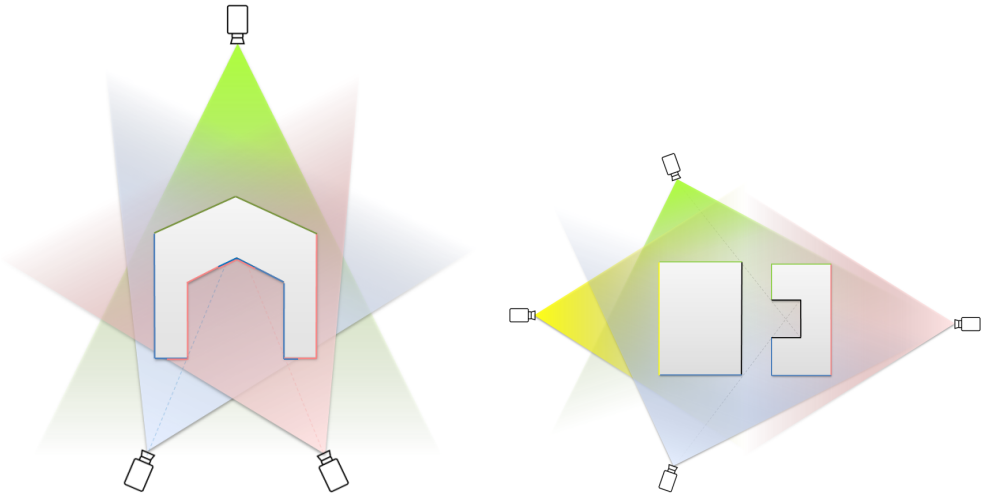


Figura 2.15: Exemplo da visualização de um objeto por três câmaras. Na figura à esquerda toda a superfície do objeto é capturada (o que era impossível de fazer com apenas dois sensores) enquanto que na imagem à direita isso continua impossível de concretizar.

que estejam em movimento, uma vez que permite tirar um "foto" ao espaço num determinado momento ou até a gravação contínua de vídeo 3D.

Já existem alguns produtos e serviços no mercado que oferecem este tipo de aquisição. A *SpaceVision*[31] tem um produto de nome *Cartesia Series Portable 3D Body Scanner* que afirma ser o primeiro sistema de *scan* corporal portátil do mundo. É constituído por nove câmaras, distribuídas igualmente por três torres, e usa *laser scanning* para realizar a aquisição de informação. A captura demora aproximadamente dois segundos a ser feita e é gerada uma nuvem de pontos com cerca de 1 milhão de pontos e com um erro de medição médio inferior a 3mm.

Figura 2.16: Cartesia Series Portable 3D Body Scanner *da* SpaceVision[30].

A *4DDynamics* comercializa um sistema para fazer a captura 3D do corpo

[30]Retirada de http://www.space-vision.jp/EP-Body_Scanner.html (acedido em outubro de 2014)

[31]http://en.space-vision.jp/ (acedido em outubro de 2014)

de uma pessoa e, como tal, está orientado para a indústria médica ou cinematográfica (Figura 2.17). A configuração é constituída por 2 a 8 *scanners* (câmaras) e é usado um sistema de projeção de padrões de luz estruturada. A captura não é feita instantaneamente (demora 1 a 2 segundos dependendo da configuração), no entanto desenvolveram uma técnica ("*steady scan*") para compensar pequenos movimentos, como aqueles resultantes do batimento cardíaco ou da respiração.

Figura 2.17: Sistema de Body Scanning *da* 4DDynamics[32].

A *IR-Entertainment*[33] utiliza um sistema em forma de anel com câmaras *Canon DSLR* de $18MP$ (Figura 2.22). A geração dos modelos é conseguida através de técnicas de fotometria e luz estruturadas, dando origem a modelos de elevada qualidade. O número de câmaras utilizadas depende do fim pretendido, variando entre as 52 para a realização de captura facial, e as 115 para a captura corporal. Em ambos os casos são produzidos modelos 360° *Gigapixel*. Este sistema consegue realizar a captura instantânea de poses estáticas ou em movimento, porém, não permite captura contínua de informação.

Figura 2.18: Sistema de captura 360° da IR-Entertainment[34].

[32]Retirada de http://www.4ddynamics.com/3d-scanners/bodyscanner/ (acedido em outubro de 2014)

[33]http://ir-ltd.net/ (acedido em outubro de 2014)

[34]Retirada de http://ir-ltd.net/ (acedido em outubro de 2014)

A $[TC]^2$ [35] tem também um sistema orientado à captura 3D do corpo humano, mas esta recorre à utilização de *Kinects* para o fazer (Figura 2.19). São utilizados 16 sensores de profundidade, espalhados equitativamente por quatro colunas. A aquisição demora cerca de 7 segundos a ser realizada e consegue atingir uma precisão de 3mm. Este sistema foi construído a pensar principalmente no mercado têxtil, podendo ser usada como cabine para a prova virtual de vestuário, ou para a extração das medidas do corpo e posterior customização e sugestão de tamanhos de roupa.

Figura 2.19: Sistema de captura 360° utilizando Kinects *da* $[TC]^2$ [36].

Existem ainda outros sistemas, estes numa vertente mais académica, capazes de realizar o mesmo tipo de captura. Em [Boehm, 2012] foram usadas técnicas muito similares às usadas pela $[TC]^2$. São utilizadas 8 *Asus Xtion Pro Live* para realizar a captura que também estão espalhadas por 4 barras verticais. Os resultados mostraram um erro máximo inferior a 20mm. Em [Alexiadis et al., 2013] a técnica usada também foi muito semelhante pelo que neste sistema forma utilizadas 5 *Kinects*, uma a cerca de 1,30m do chão apontada para cima de forma a capturar com mais detalhe a parte superior da pessoa, e as outras 4, em forma de círculo com 3,60m de diâmetro e a 1,80m de altura, de forma a conseguir detetar todo o corpo. O detalhe deste sistema não foi especificado nos resultados.

No artigo [M. Marcon, 2009] , foi desenvolvido um sistema capaz de realizar a captura facial usando apenas uma câmara e dois espelhos de forma a gerar mais perspetivas sobre o indivíduo. A partir das três perspetivas adquiridas o mapa de profundidade é construído por um sistema baseado na estereoscopia e na minimização de energia, atingindo um detalhe inferior a 4% do tamanho do modelo capturado. Já em [Lanman et al., 2007] , o sistema construído, apesar de usar igualmente uma câmara e dois espelhos, coloca-os de forma a conseguir-se extrair, numa só imagem, 5 perspetivas diferentes do objeto a capturar: uma perspetiva correspondente à visão da câmara, duas outras vindas da reflecção direta dos espelhos, e ainda mais duas resultantes da reflecção entre espelhos como se pode ver na Figura 2.20.

Este sistema utiliza o código de *Gray* como padrão de luz estruturada

[35] http://tc2.com/ (acedido em outubro de 2014)
[36] Retirada de http://tc2.com/index_3dbodyscan.html (acedido em outubro de 2014)
[37] Retirada de [Lanman et al., 2007]

Figura 2.20: Exemplos de captura 360° utilizando espelhos[37].

para fazer a aquisição 3D em tempo real. Não foram fornecidos dados sobre o tempo específico da captura nem da qualidade da mesma além de imagens representativas.

2.2.2 Aquisição móvel

Este tipo de aquisição tem mais liberdade que a aquisição estática uma vez que, tanto o sistema de aquisição como o objeto, deixa de ter a restrição de imobilidade.

O fato de se poder mover o sistema de captura ou o objeto em questão faz com que se consiga obter informação de todas as perspetivas do objeto usando apenas um sensor, limitando assim o custo do sistema. Para realizar esta tarefa basta recolher amostras das várias perspetivas, movendo, para isso, a câmara em torno do objeto ou rodando o objeto à frente da câmara. A junção das nuvens de pontos das várias perspetivas pode ser feita após as capturas ou, como em [Kainz et al., 2012] e [Newcombe et al., 2011], à medida que esta é feita e através da técnica de SLAM (*Simultaneous Location and Mapping*) [Thrun and Leonard, 2008].

No caso de objetos estáticos, apesar de levar mais tempo a realizar a captura, isto não constitui um problema pois o objeto mantém-se inalterado. No entanto, se se quiser capturar a geometria de uma pessoa, este procedimento torna-se mais complicado uma vez que qualquer movimento durante o processo introduz erro no resultado final.

Já existem alguns produtos e serviços no mercado que oferecem este tipo de aquisição:

A *Makerbot*[38] lançou o *Digitalizer*, um *scanner* 3D orientado para objetos de pequenas dimensões (20,3cm de diâmetro por 20,3cm de altura). Este sistema utiliza dois *lasers* e uma câmara de 1,3mp para realizar a captura, conseguindo produzir modelos com uma resolução de 0,5mm e um erro aproximado de 2mm. O modelo é colocado no centro do sistema, que está equipado com uma plataforma rotativa, que faz rodar o objeto de forma a realizar a captura 360°. Devido a este processo, o tempo de captura é de aproximadamente 12 minutos. Um produto muito similar é o *scanner* 3D da *Matterform* que é capaz de capturar a geometria de objetos com dimensões até 18cm de diâmetro por 25cm de altura. Vem igualmente equipado com 2 lasers e uma câmara *HD* e o detalhe dos modelos capturados é o mesmo. A velocidade de captura pode variar entre os 5 e os 10 minutos, consoante o nível de detalhe desejado. Modelos mais detalhados, mais tempo de aquisição.

Figura 2.21: Exemplos de sistemas de captura 360° móveis de pequena dimensão. À esquerda o Digitalizer *da* Makerbot *e à direita o* scanner *3D da* Matterform[39].

Um tipo de sistemas diferentes são os *scanners* de mão ou portáteis. Estes sistemas são geralmente pequenos e de fácil transporte e manuseamento. A captura é feita através da movimentação do *scanner* à volta do objeto de forma a capturá-lo de todas as perspetivas necessárias. A tecnologia utilizada por estes sistemas varia, sendo que a mais utilizada é a luz estruturada. A *Go!SCAN 3D*[40] e a *Artec MTH*[41] , por exemplo, utilizam um padrão de luz

[38] www.makerbot.com (acedido em outubro de 2014)

[39] Várias imagens. Da esquerda para a direira, retiradas de: https://store.makerbot.com/digitizer (acedido em outubro de 2014), https://matterandform.net/scanner (acedido em outubro de 2014).

[40] http://www.goscan3d.com/ (acedido em outubro de 2014)

[41] http://scan.laserdesign.com/artec-portable-scanners (acedido em outubro de 2014)

branca codificado através de uma LED, respetivamente, e apenas uma luz de um flash normal. Ambos os sistemas são utilizados a distâncias curtas (0,4m - 1m) e oferecem resultados muito precisos, atingindo uma resolução de 0,5mm e detalhes de 0,1mm.

Figura 2.22: Exemplos de sistemas de captura 360° de mão. À esquerda o scanner *da* Go!SCAN 3D *e à direita o* scanner *da* Artec MTH[42].

Outro produto que pode ser enquadrado nesta categoria é a *Kinect*. Apesar de este sensor poder ser utilizado de forma estacionária, para a realização de uma captura 360°, uma das abordagens mais comuns é a movimentação da câmara em torno do objeto de forma a realizar toda a captura. Como já foi descrito anteriormente, este sensor utiliza um padrão de luz infravermelha estruturada e permite a captura simultânea de informação de cor e de profundidade. Apesar de não ser tão precisa como outros produtos, consegue produzir resultados com uma resolução espacial de 0,75mm e detalhes de 1,5mm a um baixo custo. Várias empresas desenvolveram *software* que utiliza a *Kinect*como sensor de aquisição. Alguns exemplos são a *KScan3D*[43] ou a *ReconstructMe*[44] que, utilizando um ou mais destes sensores, permitem a aquisição de geometria de objetos singulares, ou até a reconstrução de espaços completos, como uma sala de estar ou até toda uma casa.

Já em [Molkenstruck et al., 2008], um outro tipo de sistema foi implementado utilizado espelhos de forma a aumentar o número de perspetivas a

[42] Várias imagens. Da esquerda para a direira, retiradas de: http://www.goscan3d.com/ (acedido em outubro de 2014), http://scan.laserdesign.com/artec-portable-scanners (acedido em outubro de 2014).

[43] http://www.kscan3d.com/ (acedido em outubro de 2014)

[44] http://reconstructme.net/ (acedido em outubro de 2014)

capturar. Neste caso, foi construída uma cabine utilizando dois espelhos planos e, para a aquisição, é usada pelo menos uma câmara, dando assim ao sistema três perspetivas diferentes do indivíduo para a geração do mapa de profundidade (Figura 2.23).

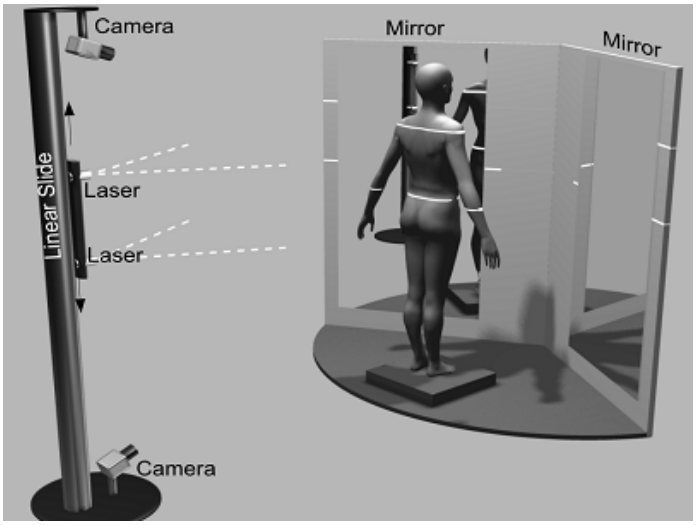


Figura 2.23: Exemplificação de uma configuração para captura 360° utilizando um sistema móvel de lasers e dois espelhos[45].

A técnica usada para a geração do mapa de profundidade foi a de *laser scanning* com um módulo móvel de *lasers* (o número de *lasers* pode variar) que se reposiciona linearmente na vertical. O tempo de geração do modelo não é especificado, no entanto, como o sistema é móvel, a captura não pode ser feita em tempo real. O erro medido neste sistema é inferior a 4mm.

2.3 Aplicações

A aquisição de informação em 360° permite capturar a geometria de um objeto de forma semiautomática, o que leva à poupança de tempo, dinheiro e materiais. Esta característica possibilitou que a criação de modelos 3D deixasse de ser exclusiva de modeladores e artistas 3D, e passasse a estar disponível para a maior parte dos públicos. Desta forma, o número de aplicações que começaram a usar este tipo de informação e a variedade das áreas aplicáveis aumentou, como já se foi mostrando nas secções anteriores. Nesta secção serão mostradas algumas das possíveis aplicações dos resultados destes *scanners*, dividindo-os nas três áreas que atualmente têm mais utilização: objetos, corpo humano e o cenário completo.

[45]Retirado de [Molkenstruck et al., 2008].

2.3.1 Objetos

A aquisição da geometria de objetos por *scanners* 3D permite que o modelo dos mesmos seja adquirido de forma muito mais facil e rápida. Dependendo dos materiais dos objetos e da técnica utilizada, consegue-se chegar a resoluções e precisões inferiores a milímetros. Esta informação pode ser usada de várias formas.

Em áreas como a indústria de videojogos e a indústria cinematográfica, estes modelos podem ser introduzidos diretamente nos processos de modelação, ou até de *render*, poupando tempo na criação dos mesmos. Além das propriedades geométricas, também podem ser recolhidas informações sobre materiais e textura, permitindo posteriormente a correta iluminação dos modelos. Mais orientada à área do consumo e aliado ao aparecimento das impressoras 3D, a aquisição de modelos e posterior replicação é também algo a ter em atenção. Isto pode facilitar processos de reparação/substituição de peças danificadas ou, simplesmente, pode ser usado para criação de objetos em miniatura de entidades reais. Apesar deste tipo de impressoras estar onde estavam as impressoras normais há 25-30 anos, as melhorias na qualidade de impressão, aumentos de desempenho e a redução de custos são vistos com frequência sugerindo assim bons indicadores para o futuro destes dispositivos.

Figura 2.24: Modelo 3D da estátua de Michelangelo[46].

Noutras áreas, como a preservação histórica, a existência dos modelos 3D de objetos históricos é também uma mais-valia, permitindo uma análise detalhada dos mesmos sem restrições espaciais nem com o perigo de danificar o objeto. Por outro lado, no caso de objetos já danificados, a aquisição 3D dos fragmentos pode ser usada também para a reconstrução digital do objeto original sem o perigo de incorrer em mais danos. Para além do estudo e visualização virtual da informação, estes modelos também podem ser usados para produzir réplicas, sejam elas miniaturas para *souvenires*, ou peças idênticas para, por exemplo, exposição temporária enquanto o original é restaurado. Já existem vários casos nesta área como, por exemplo, a estátua de *Michelangelo* (Figura 2.24) que foi captura com um detalhe de 0,29mm resultando numa malha crua de 2 biliões de polígonos.

[46]Retirada de [Levoy et al., 2000].

[47]Retirada de `http://www.makehuman.org/blog/new_realistic_teeth_model.html` (acedido em outubro de 2014)

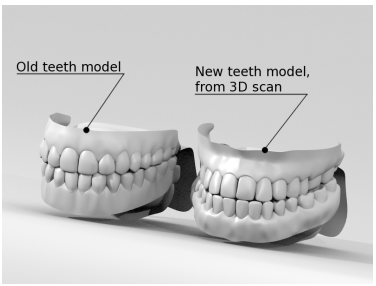


Figura 2.25: Modelo 3D da uma prótese dentária[47].

Outras áreas onde o detalhe é também importante são a indústria e a medicina. No caso da indústria, estes modelos são usados tanto para controlo de qualidade, como para processos de *reverse engineering*. Além da automatização do processo do controlo, o facto de estas atividades poderem ser realizadas por máquinas, permite que estas possam ser efetuadas em zonas inacessíveis ou perigosas para o Homem. Já na medicina, os *scanners* 3D são usados para digitalizar, por exemplo, estruturas dentárias (Figura 2.25) ou até dos ossos que depois são usadas para a geração de modelos à medida de implantes ou próteses.

Outra utilização dos modelos 3D de objetos que, de certa forma, é transversal a todas as áreas, é a prototipagem rápida. Com estes instrumentos é possível digitalizar de forma rápida informação tridimensional de protótipos construídos à mão o que permite depois ilustrar ideias com a adição de informação digital. Esta solução é cada vez mais usada por designers e pode ser útil para a criação e disseminação de ideias de novos produtos.

2.3.2 Corpo humano

A aquisição do corpo humano é uma das vertentes mais em popular dos *scanners* 3D. Além da extração do modelo de pessoas, a utilização de informação 3D para produzir informação lógica sobre o corpo humano abriu novas portas à área de interação homem-máquina. No campo dos jogos, o campo da interação natural tem como exemplo de maior sucesso a *Kinect* para a *Xbox360*. No entanto, esta não é a única área onde este tipo de interação natural é usado. A utilização de gestos é usada para navegar em menus aplicados às mais diversas áreas e, por exemplo, na robótica, indústria ou medicina, pode ser usado para controlar os movimentos de máquinas, o que pode ser útil em ambientes em que seja impossível ou perigosa a presença humana. Neste caso, mas aplicado à área da saúde, o controlo de materiais médicos sem a necessidade do contacto direto, permite inclusive a realização de cirurgias à distância com a adição de tecnologia que permita a visualização em tempo real. Ainda na área da saúde, outra utilização do reconhecimento de gestos é a interpretação de linguagem gestual. Já na área de Arte Digital, várias *performances* artísticas e peças de teatro utilizam esta tecnologia tirando partido do reconhecimento de gestos para interação com a peça.

[48]Retirada de http://kinectic.net/motion-capture-face/ (acedido em outubro de 2014)

Figura 2.26: Captura facial para o filme Avatar[48].

Por outro lado, a aquisição da geometria do corpo humano e criação do respetivo modelo, tem também várias aplicações úteis noutras àreas. Uma das mais comuns é na área do cinema e dos videojogos para a criação e animação de personagens virtuais de forma realista. No caso específico da aquisição facial, a produção de informação altamente detalhada permite animar e dar expressões naturais às personagens, como foi usado em filmes como *Timtim* e *Avatar* (Figura 2.26), ou em jogos como *Mass Effect* ou *Dead Island*. Outra área onde a geometria do corpo é também usada é no comércio a retalho, por exemplo, com a construção de provadores virtuais. Através da captura de informação 3D são extraídas e usadas as medidas e volumetria das pessoas. Com esta informação, os sistemas conseguem calcular quais as dimensões reais das pessoas e os tamanhos correspondentes que, posteriormente, através de Realidade Aumentada, podem ver como as peças de vestuário ficariam, mesmo que os artigos não estejam disponíveis no local. Como foi referido em relação à interação, a utilização apenas da geometria do corpo e a sua movimentação é também usada em peças de arte digital para interação com elementos virtuais.

Os campos da saúde e da medicina são talvez as áreas onde a utilização da geometria do corpo humano é mais útil e versátil. A aquisição da informação 3D do corpo é utilizada para monotorização e para a procura e visualização de anomalias como malformações ou resultados de acidentes. Os exemplos mais comuns são os raio-x, a tomografia computorizada e as ressonâncias magnéticas. No caso de sistemas que conseguem fazer este tipo de observações em tempo real é ainda possível a deteção de movimentos em escalas milimétricas, por vezes impercetíveis ao olho humano, que são analisados de forma a detetar irregularidades. Ainda relacionado com esse tipo de movimentos, no campo das ciências forenses, há sistemas de aquisição facial de alta precisão que permitem a deteção de micro-expressões faciais que depois são utilizadas, por exemplo, na análise de depoimentos para o apuramento da veracidade das afirmações. Outra utilização destas tecnologias na saúde é na criação de modelos a partir de partes do corpo da pessoa. Estes modelos podem ser usados para a criação de próteses ou implantes que são usados nas próprias pessoas, aumentando assim o realismo dessas peças e semelhanças com as partes originais. Esta técnica também é usada para planeamento de cirurgias plásticas. A nível do ensino, o *scan* do corpo humano, além de permitir o estudo virtual minucioso, possibilita a criação de modelos ainda mais detalhados que são usados para,

por exemplo, treinar cirurgias sem colocar em perigo nenhum paciente.

2.3.3 Ambientes

Os *scanners* 3D também podem ser usados para fazer a captura de todo a estrutura de um determinado ambiente e não apenas partes dele. Estas capturas são, por norma, dinâmicas e envolvem algoritmos simultâneos de localização e mapeamento (*SLAM*)[Thrun and Leonard, 2008].

Na área do entretenimento este tipo de captura permite a criação de ambientes virtuais semelhantes aos reais, de forma fácil e relativamente rápida. Isto pode aumentar a imersividade no campo dos jogos, uma vez que a navegação passa a ser feita num cenário virtual muito semelhante ao real. Esta característica também pode ser usada na área da cultura, possibilitando a exploração e visitas virtuais realistas de museus ou espaços que estão geograficamente espalhados pelo mundo. Já no cinema, a construção destes mundos virtuais permite a realização de "*free view-point videos*", isto é, películas em que o utilizador pode navegar pelo espaço (virtual) enquanto a narrativa ou as animações decorrem. Outro tipo de aquisição que atualmente é usada no cinema é a estereoscopia, produzindo a noção de profundidade que é mostrada nos filmes 3D.

Por outro lado, e numa vertente menos interativa, a área de preservação histórica também pode beneficiar da aquisição 3D de espaços históricos. A existência dos modelos 3D destes espaços possibilita que informação detalhada possa ser difundida digitalmente e, desta forma, chegar a muitas mais pessoas. Além disso, permite a visualização do espaço quando estes estão inacessíveis, potenciando assim a preservação dos mesmos.

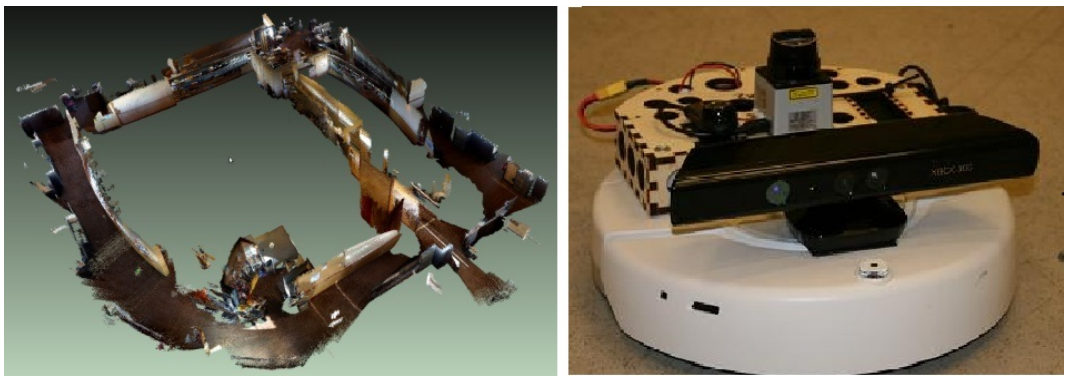


Figura 2.27: Captura de uma cena usando SLAM[49].

[49]Retirada de [Henry et al., 2010]

Outra área que também tira partido dos *scanners* 3D, principalmente daqueles aptos para a captura em tempo real, é a robótica. Atualmente existem vários projetos que integram estas duas componentes adicionando a informação tridimensional do espaço à máquina. Por um lado, isto pode ser usado para o *robot* conseguir interagir com o cenário que o rodeia, mesmo que estes sejam dinâmicos, dando-lhe a capacidade de, por exemplo, manipular objetos ou interagir com outras máquinas. Por outro, e no caso dos robots com mobilidade, a noção da tridimensionalidade do espaço facilita a navegação no mesmo através de *SLAM* (Figura 2.27). Além da possibilidade de fazer a construção do modelo desses espaços autonomamente, isto abre portas à navegação autónoma. Esta área tem tido vários desenvolvimentos nos últimos anos e é aliciante, sobretudo quando realizada ao ar livre, possibilitando avanços no campo dos automóveis autónomos.

2.4 Sumário

A captura de informação 3D está em rápida evolução e os sensores aptos para o fazerem estão a chegar aos consumidores com preços mais acessíveis pelo que, um dos sensores em maior destaque é a *Microsoft Kinect*, tanto pela sua velocidade e precisão, como pelo seu custo reduzido.

Dependendo dos objetivos pretendidos, existem várias tecnologias capazes de realizar a captura de informação em 3D com características diferentes. As que mais se destacam são a Estereoscopia, *Time-of-Flight* e Luz Estruturada. No caso dos sistemas de aquisição de 360°, existem vários mas com características diferentes que oscilam entre a qualidade e velocidade da aquisição.

As suas aplicações são variadas abrangendo áreas tão diferentes como os videojogos, indústria, medicina ou artes.

Capítulo 3

Visão do Sistema

Como foi descrito na introdução, o objetivo desta tese consiste em criar um sistema de baixo custo capaz de fazer a aquisição 360° de informação 3D de um objeto em tempo real, a partir de uma configuração estática e utilizando apenas uma câmara. Estas características geralmente contrapõem-se e, como tal, é um desafio conseguir conciliá-las no mesmo sistema.

Com isto em mente, neste capítulo pretende-se mostrar como foi construída a visão deste sistema a partir dos requisitos enunciados. Inicialmente serão descritos em detalhe os objetivos e as características pretendidas e, de seguida, serão expostas e justificadas as decisões tomadas, tanto a nível de material utilizado para a realização da captura como para a construção da configuração física. Isto determinará a forma como o sistema será construído e quais as características técnicas do mesmo. Este capítulo será concluído com a apresentação de algumas das possíveis aplicações deste sistema.

3.1 Descrição e restrições

Características como baixo-custo, aquisição 360° ou tempo real, que estão na base dos requisitos para a construção desta aplicação, não são por norma compatíveis, no entanto são características necessárias para cumprir os objetivos propostos.

A aquisição 360° de um objeto é útil para obter a geometria quase total do mesmo, apenas as partes ocludidas às perspetivas capturadas é que não são adquiridas. Este tipo de informação pode ser utilizada para diferentes propósitos como será detalhado na Secção 3.4. Esta aquisição pode ser feita de forma faseada, isto é, uma perspetiva de cada vez, ou então em simultâneo, capturando todas as perspetivas de uma só vez.

Para um método faseado conseguir realizar uma captura 360° em tempo

real seria necessário que cada uma das fases fosse suficientemente rápida para conseguir capturar todas as perspetivas do objeto numa fração de tempo muito pequena. Assumindo uma taxa de atualização de 30 *frames* por segundo e considerando apenas 3 perspetivas distintas, seria necessário que cada fase tivesse um tempo de 1/90 segundos. Para conseguir isso seria necessário ou um elevado número de dispositivos de aquisição ou uma estrutura que permitisse rodar o objeto (se possível) ou a câmara a velocidades elevadas. Neste último caso, a câmara também teria de ter um tempo de exposição mínimo, de forma a conseguir capturar todas as perspetivas consecutivamente. Estas restrições fariam com que o custo do sistema fosse muito elevado e, como tal, esta abordagem foi descartada.

O requisito de utilização de uma única câmara está apenas relacionada com o fato de se querer que o sistema também seja de baixo custo. Apesar de apresentar uma dificuldade adicional na conceção do sistema, este é um fator a ter em conta já que os sensores de aquisição, por norma, não são acessíveis. Desta forma, a utilização de múltiplos sensores é descartada uma vez que, além de aumentarem a carga computacional do sistema, também aumentam o seu custo.

Já existem algumas soluções que permitem fazer a aquisição 360° de informação 3D, no entanto, estas têm algumas características diferentes das propostas nesta dissertação. As soluções da *4DDynamics* e a *IR-Entertainment* funcionam como estúdios e estão orientadas a modelos com grandes dimensões e à aquisição de modelos estáticos. Isto faz com que estas soluções tenham um nível de detalhe muito elevado mas em contrapartida requerem instalações de grande dimensão que, devido à qualidade e quantidade de material utilizado, são também soluções dispendiosas. A *KScan3D* por sua vez oferece uma solução mais acessível, utilizando *Kinects* como instrumento de captura, e permite realizar a aquisição 360° de objetos estáticos mas de forma faseada. Os modelos obtidos têm também menos pormenor que as soluções anteriores devido às menores capacidades do *hardware* usado.

Outras soluções, como o *Digitalizer* ou o *scanner* 3D da *Matterform*, têm objetivos diferentes e estão orientadas à portabilidade e à comercialização para o público geral. Estes *scanners* conseguem fazer a aquisição 3D a 360° de objetos estáticos de pequena dimensão. Estão pensados para a criação de modelos completos e detalhados e, pensando nas impressoras 3D, a sua replicação.

3.2 Decisões

No que toca à perceção do conteúdo de uma determinada cena, os computadores têm algumas limitações uma vez que esta é, por norma, representada com

informação em duas dimensões. Isto gera alguns problemas em campos como a segmentação da cena e a deteção e reconhecimento de objetos. A introdução da tridimensionalidade pode ajudar na resolução de algumas dessas questões. De acordo com as características pretendidas e os objetivos traçados, os principais pontos de decisão prendem-se com o dispositivo utilizado para a realização da captura de informação e a configuração a usar, de forma a conseguir fazer a captura 360° em tempo real.

3.2.1 Captura

A captura de informação consiste na conversão de informação do mundo real para o formato digital. No caso do vídeo e da aquisição 3D, a informação a capturar consiste na imagem e geometria das cenas. No capítulo anterior foram referidas várias abordagens para a captura de informação em 3D no entanto, nem todas servem para a realização deste projeto. Além da qualidade da informação adquirida, questões como o tempo de captura, preço e o facto de se querer um sistema de captura 360° são os principais fatores de decisão.

Os sistemas baseados em estereoscopia são frequentemente usados na indústria cinematográfica para a realização de filmes com perspetiva 3D. A qualidade dos resultados produzidos encontra-se apenas na ordem dos centímetros, no entanto é o suficiente para o tipo de aplicações pretendidas. Normalmente são utilizadas duas câmaras para efetuar a captura e produzir o efeito 3D, o que faz com que estes sistemas sejam de baixo custo e de fácil integração. Apesar da obtenção da informação em 3D exigir uma grande complexidade computacional a nível de processamento de imagem, esta pode ser processada em tempo real com os processadores mais recentes. Em contrapartida, a informação produzida está dependente do ponto de vista a partir do qual foi capturada, pelo que, para aquisição 360° torna esta técnica pouco viável. A aquisição teria de ser efetuada de forma faseada, perdendo a característica de tempo real, ou utilizando múltiplas fontes e captura dispersas pelas várias perspetivas, o que faria com o que o custo da aplicação aumentasse.

Por outro lado, os sistemas *Time-of-Flight* são mais versáteis e podem ser usados para medições de curto a longo alcance nas mais diversas áreas. Estes sistemas são reflexivos e ativos, isto é, emitem radiação para conseguir fazer as medições, através do reflexo das mesmas no objeto. Apesar de ser uma tecnologia que também exige uma carga computacional elevada, os sensores começam agora a ter a capacidade de fazer esse processamento localmente, diminuindo assim a complexidade do *software* que a usam. A qualidade dos resultados pode chegar a precisões na ordem dos milímetros, dependendo das condições, e têm uma taxa de atualização elevada, tornando-os úteis, por exemplo, para sistemas interativos. Da mesma forma, estes sensores geram uma quantidade

de tráfego elevada, exigindo assim interfaces capazes de o suportar. Por esta razão, e dada a necessidade de utilização de materiais emissivos como *LEDs* ou *lasers*, estes sistemas têm por norma um custo mais elevado.

Os sistemas de Luz Estruturada primam por serem sistemas capazes de produzir resultados de elevada qualidade chegando a poder atingir detalhes de micrómetros. Uma vez que utilizam a emissão e análise de padrões, esta tecnologia está mais orientada a ambientes *indoor* controlados e é ideal para a captura da geometria de objetos estáticos e criação de modelos 3D. Apesar de ser uma tecnologia computacionalmente exigente pode atingir-se taxas de atualização interativas, no entanto nem todas as variantes desta tecnologia o permitem. Os padrões sequenciais, como o código de *Gray*, necessitam da emissão de várias imagens com padrões diferentes para a aquisição de uma única cena, o que faz com que o processo demore mais tempo a ser executado. Por outro lado, o tipo de radiação emitida também pode influenciar a captura se a obtenção da informação sobre as cores do objeto for um dos objetivos pretendidos. Se o emissor utilizar padrões de luz do espetro visível, estas podem adulterar a aquisição dessa informação. Uma forma de contornar esses casos pode passar por usar padrões do espetro infravermelho, tal como é feito com as câmaras de profundidade como a *Kinect*. Uma vez que os sistemas de Luz Estruturada são sistemas ativos e necessitam também de um sistema de projeção, dependendo da escolha, pode fazer com que o preço dos sistemas seja mais elevado.

Das diferentes metodologias analisadas, aquela que mais se adapta às características do sistema proposto é a de Luz Estruturada devido ao detalhe das capturas e ao tempo necessário para as efetuar. Dos vários sistemas existentes que utilizam esta tecnologia, aqueles que mostraram ter um melhor compromisso entre qualidade, *performance* e custo foram as câmaras de profundidade. Estes sensores têm como principal vantagem o acesso explícito à informação 3D com boa qualidade e de forma rápida. Além disso, o baixo custo e a portabilidade são também fatores que os valorizam. Em contrapartida, os sensores de profundidade apresentam desvantagens como o alcance limitado, tanto a distâncias muito curtas (0-50cm) como longas (superiores a 5 metros), e dificuldade na aquisição de informação em condições menos favoráveis, como a exposição direta de luz solar nas superfícies ou a aquisição de materiais brilhantes/refletores ou transparentes.

Dentro desta gama de sensores, foram estudadas as características de quatro dispositivos presentes no mercado: a *Kinect*, a *Xtion PRO Live* e as *Carmine* 1.09 e 1.08, como se pode ver no seguinte quadro.

Todas elas apresentam características muito semelhantes em relação às suas especificações e custos de aquisição. Apenas a *Carmine* 1.09 se diferencia um pouco uma vez que está orientada apenas para aquisições de curto alcance. En-

	Kinect / Kinect4Windows	Xtion PRO LIVE	Carmine 1.08	Carmine 1.09
Distância Mínima	0,8m/0,4m	0,8m	0,8m	0,35m
Distância Máxima	3,5m	3,5m	3,5m	1,40m
FOV Vertical	43°	45°	45°	45°
FOV Horizontal	57°	58°	57,5°	57,5°
Imagem de Profundidade	640×480 (30*fps*)	640×480 (30*fps*)	640×480 (60*fps*)	640×480 (60*fps*)
Imagem de Cor	1280×960 (10*fps*) 640×480 (30*fps*)	1280×1024 (15*fps*)	640×480 (30*fps*)	640×480 (30*fps*)
Preço	€150 / €250	£139 (€166)	$200 (€148)	$200 (€148)

Tabela 3.1: Comparação das características das câmaras RGBD Kinect, Xtion PRO Live, Carmine *1.09 e 1.08. Os valores de distância apresentados são nominais, sendo possível no entanto operar fora destes intervalos.*

tre as quatro câmaras, decidiu-se utilizar a *Kinect* uma vez que é o dispositivo mais popular e com mais suporte por parte da comunidade de desenvolvedores. Das duas soluções que a *Microsoft* apresenta, a escolha recairia pela *Kinect* para a *Xbox* uma vez que é mais económica e as características são muito semelhantes. No entanto, e para os testes realizados, o sensor utilizado foi a *Kinect for Windows*, cedida pelo CCG[1], tendo como única vantagem a maior precisão para capturas efetuadas a distâncias inferiores a 80 cm da câmara.

3.2.2 Configuração 360°

A captura 360° de informação pressupõe que esta seja adquirida de várias perspetivas. Uma das formas de o fazer é através da movimentação do sensor à volta do objeto ou pela movimentação do objeto em frente do sensor, tal como é utilizado nos sistemas da *Go!SCAN 3D* ou no *Digitalizer* respetivamente. No entanto, devido à característica pretendida de tempo real, a captura faseada destas diferentes perspetivas não é viável a custos suportáveis. Outra das formas de efetuar este tipo de captura é através da utilização de várias fontes de captura. Esta abordagem é usada em vários sistemas como o da *IR-Entertainment* ou no sistema referido em [Alexiadis et al., 2013], no entanto tem como desvantagens um maior custo em *hardware* e também o aumento da carga computacional.

No caso específico da utilização de várias *Kinects*, além dessas questões existe ainda outro problema relacionado com a interferência causada entre estas. Quando, por exemplo, duas *Kinects* estão direcionadas para uma determinada zona existem áreas de sobreposição, isto é, zonas cuja aquisição é feita por ambos os sensores. Nestas regiões há a formação de ruído devido à sobreposição dos padrões *IR* que impossibilitam a correta estimativa da profundidade dessas áreas (Figura 3.1). Nos testes realizados verificou-se ainda

[1] Centro de Computação Gráfica (www.ccg.pt)

que existe mais ruído quando as câmaras se encontram em posições frontais.

Figura 3.1: Representação do ruído causado pela interferência entre duas Kinects.

Outra forma de se conseguir capturar informação de várias perspetivas de um objeto é utilizando espelhos. Devido às suas propriedades refletoras, este material permite a visualização de informação de diferentes perspetivas a partir de um único ponto de vista. A utilização de espelhos para a realização de captura em 3D já foi concretizada noutros sistemas de aquisição como em [Alexiadis et al., 2013] ou em [Molkenstruck et al., 2008]. A colocação do espelho em frente ao sensor e por trás do objeto permite que seja capturada informação de várias perspetivas com apenas um sensor. Dependendo do número de espelhos utilizados e da posição relativa entre eles é ainda possível aumentar o número de perspetivas diferentes, tal como foi demonstrado em [Lanman et al., 2007].

No entanto, a utilização de espelhos com a *Kinect* não se encontra documentada, pelo que a única referência encontrada até à data da escrita desta dissertação encontra-se no livro *Making Things See* [Borenstein, 2012]. Aqui é mencionada a utilização de espelhos pelo artista e investigador *Kyle McDonald* como um exemplo da versatilidade da *Kinect* (Figura 3.2) mostrando como esta se comporta na presença de espelhos. No entanto, não foram encontradas evidências de este trabalho ter tido continuidade.

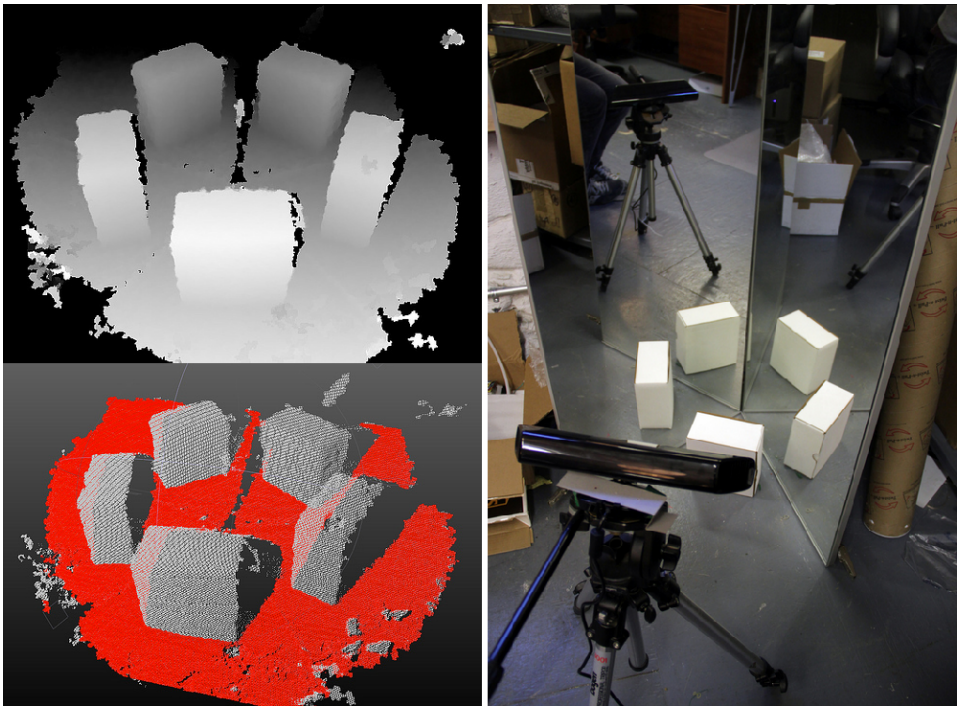


Figura 3.2: Setup de captura Kyle McDonald *utilizando uma* Kinect *e dois espelhos*[2].

De acordo com as características necessárias para a construção do sistema e as características das soluções analisadas, o melhor compromisso encontrado foi a utilização de um sistema de espelhos. O posicionamento destes materiais na periferia da cena permite recolher informação sobre o objeto de outras perspetivas como se se tratassem de pontos de aquisição virtuais. Estes são capturados também pela câmara, assim como uma visão direta do objeto, maximizando desta forma a quantidade de informação recolhida de uma só vez. Apesar de este processo continuar a exigir uma carga computacional acrescida a nível de análise de imagem, esta carga é igual ou menor à existente no caso de múltiplas câmaras. Além disso, o custo global do sistema diminui uma vez que os espelhos são menos dispendiosos que os sensores: um espelho de 80 cm por 120 cm consegue ser adquirido por 20€.

3.3 Características do sistema

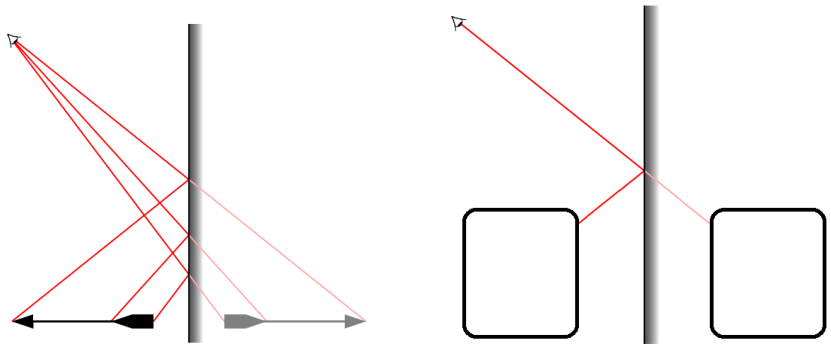
De acordo com as decisões tomadas, foi escolhido implementar um sistema que utiliza apenas um sensor de aquisição, a *Kinect*, e uma configuração de

[2]Retirada de http://www.flickr.com/photos/kylemcdonald/5641883004/ (acedido em outubro de 2014)

espelhos de forma a realizar a captura 360°. Estas decisões permitem que haja alguma liberdade quanto à estrutura física do sistema, podendo este ser disposto de diversas formas consoante os objetivos da captura. No entanto, e apesar de não serem completamente independentes da configuração utilizada, as características do sistema estão intimamente ligadas às limitações dos materiais escolhidos.

3.3.1 Espelhos

O espelho é um objeto que reflete a luz numa direção bem definida, em vez de a absorver ou refletir em todas as direções [Wikipedia, 2014]. No caso específico dos espelhos planos, que são os mais comuns e os que serão usados no sistema proposto, qualquer imagem que por eles seja refletida mantem as dimensões originais. Os espelhos são utilizados em diversas áreas como a ciência ou na indústria para a produção de lasers, câmaras e telescópios, mas a sua utilização mais comum é a doméstica. No presente caso, a utilização dos espelhos vai servir para simular pontos de vista virtuais de forma a capturar a informação do objeto numa perspetiva 360°.



(a) (b)

Figura 3.3: Demonstração da reflexão dos espelhos. À esquerda, a relação entre os ângulos de reflexão e a normal do espelho. À direita, demonstração da simetria para cálculo das distâncias das superfícies refletidas.

O objetivo destes pontos de vista virtuais é a extração de informação 3D do objeto capturado de outras perspetivas e, como tal, será necessário usar também a reflexão do padrão infravermelho da *Kinect* para o fazer. Tanto neste espetro como no espetro visível, a direção em que a luz é refletida está diretamente relacionada com o plano do espelho e a sua normal (Figura 3.3(a)). A distância de um determinado ponto que seja capturado através da sua reflecção num espelho corresponde à soma da distância desse ponto ao espelho e

da distância do espelho ao sensor. Tendo esta informação e uma vez que a reflecção é simétrica ao plano do espelho, o cálculo efetuado para a obter correta posição dos pontos refletidos (Figura 3.3(b)) resume-se à seguinte equação:

$$P = P_{mirror} + 2 \times d \times \vec{N}$$

Onde, P_{mirror} é a posição original do ponto refletido no espelho, d a distância do ponto P ao plano do espelho e $\vec{N}$ a normal desse mesmo espelho.

Os espelhos são por norma constituídos por uma camada de material refletor (prata, alumínio ou amálgama de estanho) e, por cima, uma lâmina de vidro [Wikipedia, 2014]. Uma vez que a luz passa por duas superfícies, o vidro e depois a camada refletora, este tipo de construção faz com que possa ser introduzida na imagem uma "sombra" e como tal, que seja gerado algum erro (Figura 3.4). Quanto mais espessa for a lâmina de vidro, maior o erro associado. No caso desta dissertação os espelhos utilizados têm uma espessura de aproximadamente 4mm o que faz com que o erro associável seja mínimo e, como tal, não será tomado em conta para os cálculos necessários.

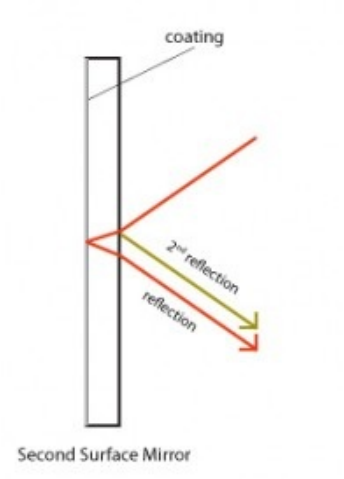


Figura 3.4: Efeito de "sombra" nos espelhos causado pela placa de vidro[3].

3.3.2 *Kinect*

Como já foi abordado no Capítulo 2, a *Kinect* é um caso de sucesso no que toca a periféricos para jogos e teve também uma grande aceitação na comunidade de desenvolvedores de *software*, dando desta forma origem a várias aplicações

[3]Retirada de http://forum.david-3d.com/viewtopic.php?p=9285a (acedido em outubro de 2014)

nas mais diversas áreas. As suas principais características são a captura de informação de profundidade de boa qualidade e de forma rápida ($30fps$), a integração simultânea de uma câmara RGB e baixo custo do material. Por outro lado, as suas principais limitações prendem-se com os limites de distâncias para captura, tanto inferior como superior, e a perda ou ruído na informação na presença de condições adversas como a exposição das superfícies à luz solar ou na captura de superfícies com propriedade reflexivas.

O método de aquisição da informação de profundidade por parte da *Kinect* baseia-se em luz estruturada, isto é, é emitido um padrão sobre a cena que depois de interpretado consegue produzir a informação de profundidade da mesma. Neste caso, o padrão emitido é um padrão ruidoso que é emitido no espetro infravermelho o que possibilita a aquisição simultânea de informação de profundidade e de cor sem que uma influencie a outra.

O mapa de profundidade produzido tem dimensões de 640×480 pixéis, uma amplitude vertical de 43°e horizontal de 57°e consegue capturar informação a distâncias que vão dos 0,5m aos 3,5m. No entanto, e como já foi anteriormente referido, é possível realizar a captura de informação a distâncias mais curtas que as tabeladas. Este mapa é representado em forma de matriz em que cada elemento tem $11bits$: um é usado para representar informação errada e os outros 10 para a detalhar a profundidade, estando assim disponibilizados 1024 níveis de profundidade para a representar [Aouina et al., 2013]. Cada um destes níveis é usado para guardar a representação de uma "fatia" do espaço, perpendicular ao eixo ótico do sensor (Figura 3.5).

Como se pode inferir a partir da figura, a distância de entre cada um destes níveis não é constante e é esta distância que reflete o erro das medições. O tamanho do passo (q) entre cada nível é dado pela seguinte função quadrática [Smisek et al., 2011]:

$$q(z) = 2,73z^2 + 0,74z - 0,58[mm]$$

Através dela é possível inferir o erro associado a uma medição de acordo com a distância a que os objetos se encontram da câmara, como se pode ver no seguinte Gráfico 3.6:

Para distâncias pequenas o passo entre níveis é pequeno, oferecendo desta forma uma boa resolução, enquanto que as medições efetuadas a objetos que se encontram mais distantes já têm um passo elevado o que deteriora a qualidade da aquisição. Além disso, medições efetuadas a distâncias muito curtas (inferiores a aproximadamente 350mm) tomam valores negativos refletindo a incapacidade da *Kinect* em realizar medições nesse intervalo espacial.

[4]Retirada de [Smisek et al., 2011]
[5]Retirada de [Smisek et al., 2011]

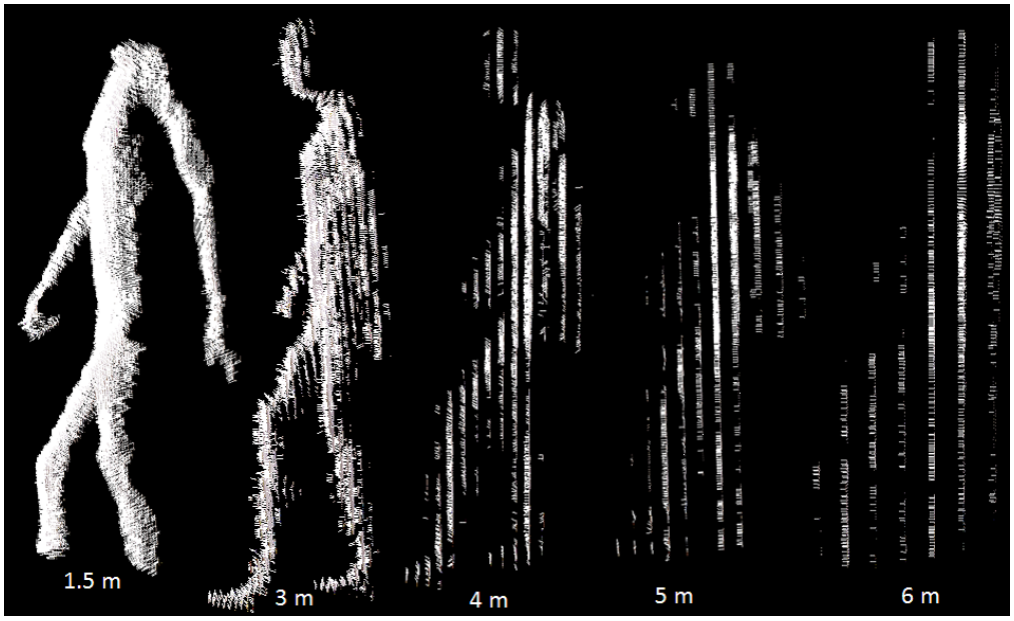
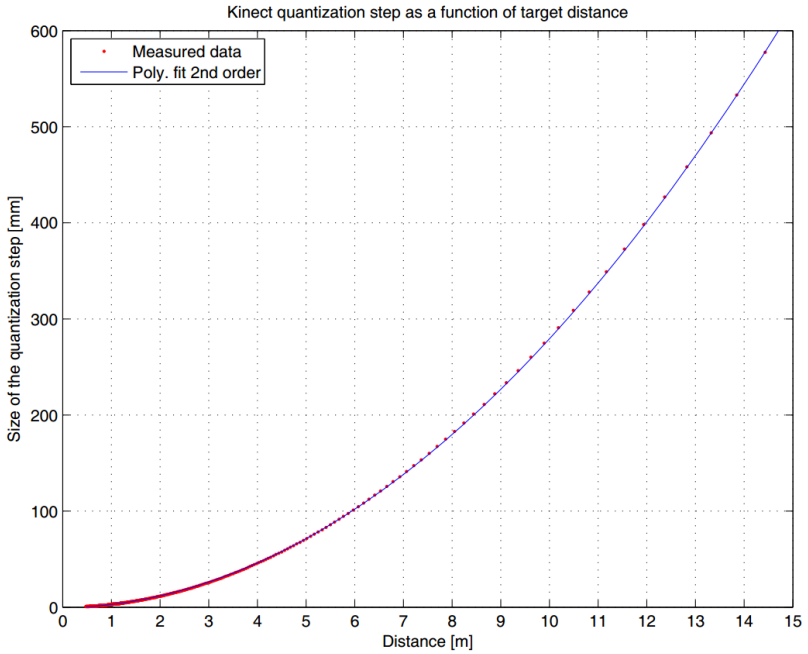


Figura 3.5: Representação da informação de profundidade da Kinect *por "fatia" espacial*[4].



(b) Kinect depth quantization step q (0-15 m).

Figura 3.6: Gráfico demonstrativo do nível de detalhe vs. *distância da informação de profundidade da* Kinect[5].

Além desse erro, a qualidade de aquisição também é influenciada por fatores externos como a exposição a outras fontes de radiação ou até pelos próprios materiais constituintes dos objetos a capturar. A *Kinect* apenas consegue capturar informação de profundidade em zonas onde o padrão emitido é visível pelo sensor, sem outras interferências maiores. Se esta informação for refletida ou simplesmente distorcida, tanto pelas propriedades do objeto como pela presença de outros padrões (como por exemplo a interferência entre padrões de várias *Kinect*s), essa área é representada com um valor de erro o que, no mapa de profundidade, é representada como zona sem informação (zonas a preto). No caso do sistema proposto, estes são os erros que mais influenciam a qualidade das medições efetuadas, pelo que a *Kinect* é o fator mais importante no resultado final.

Associando as limitações da *Kinect* à utilização dos espelhos, outro problema que surge é a restrição a nível do espaço disponível para usar. Um ponto que seja capturado pela *Kinect* através da sua reflecção num espelho encontra-se a uma distância superior à real, mais precisamente à soma da distância entre o "raio" que se estende do sensor ao espelho e do espelho ao ponto. Desta forma, o espaço disponível e a dimensão dos objetos que podem ser capturados ficam limitados e são inferiores aos da *Kinect*. No entanto, como estes limites dependem da configuração usada, não podem ser definidos de forma fixa e como tal serão abordados na Secção 4.1.

3.4 Casos de uso

Um sistema de captura de informação em 3D consegue, de forma mais ou menos eficiente, adquirir a informação geométrica de objetos ou outras entidades no formato de uma nuvem de pontos. Esta nuvem representa a entidade capturada e pode facilmente ser visualizada através de software como o *MeshLab*[6]. Uma vez criadas mais que uma nuvem de pontos, estas podem também ser unidas de forma a criar uma representação única do objeto com mais qualidade. O sistema proposto tem como objetivo a aquisição 360° da informação geométrica de um objeto (ou outras entidades) com taxas de atualização interativas, como tal, será capaz de capturar e unir a informação proveniente das nuvens de pontos das várias perspetivas em simultâneo.

Estas características fazem com que seja possível fazer a aquisição 360° da geometria de um modelo em tempo real, o que abre portas a outro tipo de aplicações para além da simples captura de modelos. No entanto, e devido às restrições dos materiais e configuração usadas, o sistema tem várias limitações: a área de ação é limitada pelo alcance da *Kinect* e pelo posicionamento dos

[6] http://meshlab.sourceforge.net/ (acedido em outubro de 2014)

espelhos e a qualidade da aquisição é também influenciada pela distância a que estes se encontram do objeto a capturar.

3.4.1 Modelos 3D de Objetos

Como já foi explicado anteriormente, a criação de modelos 3D de objetos permite gerar uma representação digital da geometria de um dado objeto físico e também capturar outras propriedades como a cor ou texturas. Além da nuvem de pontos, para este fim é também útil a geração de *mesh* do objeto, isto é, criar as relações entre os pontos da nuvem de forma a gerar uma malha poligonal representativa do mesmo. Existem vários métodos para atingir este objetivo, no entanto, este processo é por norma computacionalmente exigente e como tal, difícil de atingir em tempo real.

Utilizando o sistema proposto, a geração de modelos em 3D pode ser conseguida de forma simples. A nuvem de pontos de um determinado objeto pode ser criada a partir de uma única captura, isto é, a partir da informação de uma única *frame*, no entanto esta contém por norma algum ruído proveniente da *Kinect*. Uma vez que os modelos 3D exigem uma qualidade elevada, uma forma de contornar este problema pode passar por uma captura prolongada através da utilização de várias *frames* gerando assim a nuvem de pontos do objeto de forma incremental. Assim, esta exposição prolongada do objeto ao processo de captura permite que seja recolhida mais informação e desta forma consegue-se (através de *software*) minimizar o ruído inerente à *Kinect*. A geração de malha poligonal pode ser feita posteriormente, depois de capturada a informação da nuvem de pontos.

Estes modelos podem ser usados para vários fins pelo que a aplicação mais direta é a integração dos mesmos em mundos virtuais. Áreas como os jogos de vídeo, cinema e realidade aumentada podem beneficiar deste sistema através da integração de modelos de objetos do dia-a-dia nos seus cenários. Outra aplicação que pode ter interesse inclusive a nível doméstico é a criação de modelos de objetos tendo como objetivo a sua replicação através de impressoras 3D. Esta vertente revela-se especialmente útil no que toca à substituição de peças ou componentes danificados ou perdidos.

3.4.2 Video 3D Real

A geração 360° de informação 3D em tempo real faz com que seja possível visualizar a cena a partir de qualquer ponto de vista circundante. A gravação desta informação possibilita que se consiga obter vídeo em 3D real independente da perspetiva de aquisição. Isto permite também que possam ser geradas representações 3D de objetos animados ou seres vivos em movimento em tempo

real.

As prioridades de uma aquisição em movimento e em tempo real passam por uma elevada taxa de atualização em detrimento da qualidade da captura. Neste caso, uma vez que a informação se encontra em constante atualização, o ruído originado pela *Kinect* entre *frames* é incoerente e, uma vez que existe movimentação na cena, é pouco percetível. Este ruído apenas é mais notório em superfícies estáticas uma vez que dará a perceção que a superfície se está a mexer. Uma forma de contornar este problema é a aproximação da nuvem de pontos a modelos predefinidos das entidades a capturar. Por exemplo, se se estiver a realizar a aquisição do modelo de uma pessoa, esta pode ser aproximada ao modelo de um esqueleto humano de forma a reduzir o ruído e tentar colmatar possíveis falhas no processo da captura. O mesmo se pode fazer com modelos já existentes de objetos como mesas, cadeiras, etc.

Na vertente de vídeo, este tipo de aquisição pode dar origem a vídeos em 3D real, independentes do ponto de vista, e que por essa mesma razão permitem a navegação no espaço durante uma reprodução de uma gravação. Esta característica pode ser interessante para novas abordagens na área de *storytelling*, uma vez que pode permitir que se desenrolem várias histórias ao mesmo tempo mas em espaços diferentes. Já na vertente de aquisição e visualização tempo real, e aliado a uma ligação de dados de alta velocidade, esta tecnologia pode permitir a comunicação entre pessoas em vídeo 3D real. Partindo deste mesmo princípio mas aplicando-o a uma área mais ambiciosa, a partir de um sistema deste género estaria também disponível a informação necessária para a criação de hologramas e a comunicação entre pessoas através dos mesmos.

3.4.3 Análise interativa de modelos

A geração de modelos 3D de objetos a partir de uma nuvem de pontos é útil para a integração dos mesmos em diversas áreas. No entanto, outra forma de utilizar a informação gerada prende-se com a análise e extração de características sobre essa nuvem de pontos que permita inferir outras informações que não apenas a sua estrutura geométrica. Exemplos disso são a segmentação de informação para, por exemplo, separação de diferentes componentes ou a extração do esqueleto das estruturas das nuvens de pontos cruas, isto é, o conjunto de linhas-guia das partes mais representativas dessa nuvem.

O sistema proposto consegue gerar nuvens de pontos de um determinado objeto ou entidade em tempo real. Para gerar informações como o esqueleto da nuvem e uma vez que elas se tratam de aproximações, a extração de nuvens pode ser feita diretamente sem que para tal seja necessário processamento prévio dos dados da aquisição. Desta forma é possível analisar e gerar informação

adicional de forma interativa e aplica-la assim a vários casos de uso.

Este tipo de informação pode ser bastante útil em sistemas de realidade aumentada. Imaginando o caso de um provador de roupa virtual, a extração deste esqueleto e de outras propriedades como a altura da pessoa e a espessura dos seus membros permite com que seja possível produzir uma experiência de realidade aumentada mais rica. Noutra área, mais ligada à realidade virtual, a geração deste esqueleto também pode ser usado para a animação de personagens e como a aquisição é a 360° tem mais resistência às oclusões devido às movimentações da pessoa. A segmentação também pode ser utilizada em sistemas de Realidade Aumentada de forma a separar os objetos uns dos outros e estes do cenário. Esta compreensão do cenário faz com que a sobreposição de informação seja mais fácil de ser conseguida.

3.5 Sumário

Os objetivos desta tese passam por criar um sistema de baixo custo capaz de fazer a aquisição 360° de informação 3D de um objeto em tempo real. Estas características nem sempre são compatíveis e como tal foram analisados os melhores métodos e tecnologias para o fazer e tiveram que ser tomadas decisões em duas áreas principais: o sensor responsável pela captura de informação e a configuração usada para realizar a captura 360°.

Em relação à captura foi decidido utilizar-se o sensor da *Microsoft*, a *Kinect*. Este dispositivo tem um custo relativamente reduzido, aproximadamente 150€, e tem como principais vantagens o acesso rápido e de boa qualidade a informação de profundidade da cena. Quanto à configuração utilizada, a escolha recaiu pela utilização de uma única câmara e de espelhos para capturar as diferentes perspetivas. Desta forma o custo do sistema é o mais reduzido (em comparação com as outras alternativas) e é feita uma maximização da utilização dos recursos disponíveis. As decisões do material a usar trazem alguma versatilidade na construção do sistema no entanto também limita as dimensões dos objetos a capturar e a qualidade das aquisições. Em capturas onde o objeto se encontra a distâncias superiores a 3,5m da *Kinect*, o erro associado à medição está na ordem dos 5cm o que, em certas aplicações, pode já não ser aceitável.

Foram ainda mostradas algumas das possíveis aplicações deste sistema que recaem em áreas como a construção de modelos 3D, geração de vídeo em 3D real e transmissão do mesmo em tempo real e a análise e geração de informação adicional aos modelos, como a segmentação da cena e extração de esqueletos, de forma interativa.

Capítulo 4

Implementação

De acordo com as decisões tomadas, foi escolhido implementar um sistema que utiliza apenas um sensor de aquisição, a *Microsoft Kinect*, e uma configuração de espelhos de forma a realizar a captura 360° de uma determinada entidade. Estas decisões permitem que haja alguma liberdade quanto à estrutura física do sistema, podendo este ser disposto de diversas formas. De modo a criar um sistema genérico capaz de aceitar N espelhos na sua constituição foi necessário conceber e implementar uma arquitetura base, também ela genérica, nunca esquecendo a necessidade da realização da captura em tempo real.

Assim, neste capítulo será descrita a arquitetura lógica genérica do sistema pretendido e de seguida serão apresentados os casos de estudos implementados para demonstrar as suas potencialidades. Posteriormente, o fluxo de execução será explicado, sendo cada uma das fases identificadas descrita em detalhe. Por fim serão discutidos os problemas encontrados durante a realização dos testes e quais as soluções propostas. Este capítulo será concluído com a inclusão de uma listagem e descrição das tecnologias utilizadas para o desenvolvimento do projeto e com um resumo dos tópicos abordados.

4.1 Arquitetura do sistema

Para realizar a captura 360° de um objeto, este torna-se o "centro do mundo" pelo que todos os pontos que geram informação útil para a captura encontram-se à sua volta, apontados a si. Uma vez que se utilizará apenas uma câmara para realizar a captura de informação, é necessário adaptar a disposição dos outros componentes do sistema a esta de modo a maximizar a informação capturada.

Desta forma, o sensor estará numa posição fixa apontada para a área de captura de forma a ter uma visão frontal da mesma. Uma vez que a imagem da

Kinect tem um formato 4 : 3 (a resolução é de 640×480), a largura do campo de visão é superior à sua altura. Com isto em mente, para capturar objetos cuja largura é igual ou inferior à sua altura, a *Kinect* deve ser posicionada de forma a adaptar altura do objeto na imagem à altura da imagem capturada (Figura 4.1). Assim, as margens laterais da imagem podem ser utilizadas para adquirir informação proveniente dos espelhos, que terão ser colocados de acordo com isso.

Figura 4.1: Ajuste da posição da Kinect *de acordo com a dimensão do objeto a capturar. Dadas as dimensões do objeto, é possível ajustar a posição da câmara à altura do objeto e ter na mesma espaço para colocar os espelhos.*

Por outro lado, se o objeto a capturar tiver uma largura superior à sua altura, o ajuste referido anteriormente faz com que este passe a ocupar a quase totalidade da imagem. Neste caso, a melhor opção é a colocação da *Kinect* numa posição mais distante ao objeto de forma a criar uma margem na imagem à volta dele (Figura 4.2). Esta margem poderá também ser usada para extrair a informação adicional proveniente dos espelhos e para definir o correto posicionamento dos mesmos.

Relativamente aos espelhos, o seu posicionamento está dependente do da *Kinect* e estes deverão ser colocados numa posição oposta ao sensor, isto é, do outro lado da zona de captura, de forma a serem visíveis pelo mesmo. Além disso, os espelhos devem estar dispostos de modo a serem capturados pela imagem nas áreas que não são ocupadas pela zona de captura o que, por norma, corresponde às áreas da periferia da imagem (Figura 4.3). A orientação dos espelhos em relação ao objeto está também dependente da posição da câmara:

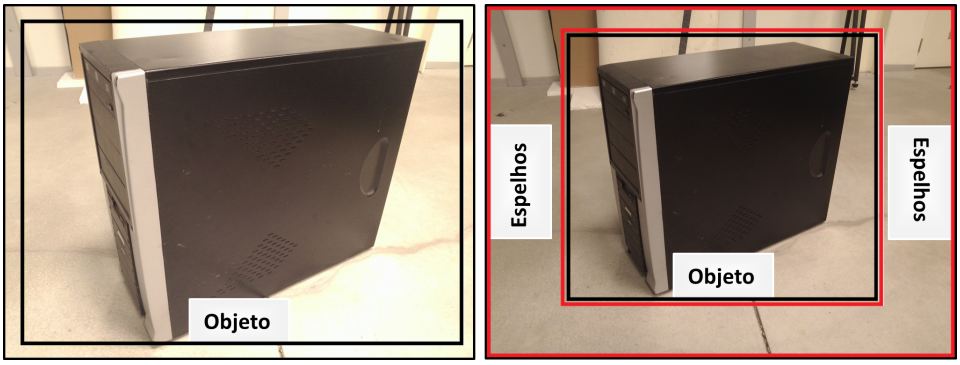


Figura 4.2: Ajuste da posição da Kinect *de acordo com a dimensão do objeto a capturar. Dadas as dimensões do objeto, é possível ajustar a posição da câmara à altura do objeto e ter na mesma espaço para colocar os espelhos.*

de forma a produzirem informação útil, estes devem ser posicionados de modo a refletirem a área de captura para a câmara.

Assim como foi utilizado em [Lanman et al., 2007] e mostrado no trabalho de *Kyle McDonald*, também foi estudada a utilização da reflexão entre espelhos de forma a adquirir ainda mais pontos de vista sobre o objeto. No entanto, esta hipótese foi descartada uma vez que introduziria uma maior complexidade ao desenvolvimento do mesmo e, além disso, adicionaria uma restrição adicional relativamente ao espaço disponível para a captura. Uma vez que a *Kinect*

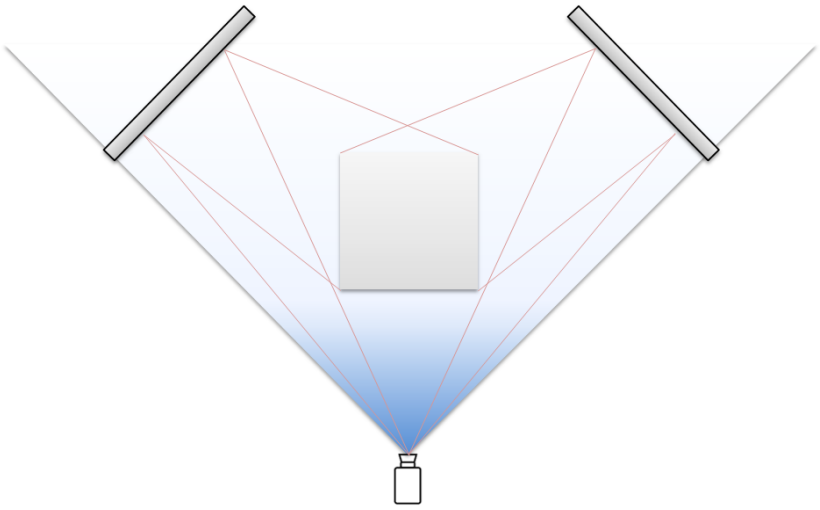


Figura 4.3: Esquema em vista de topo do posicionamento dos espelhos em relação à posição da Kinect *e da área de captura. Os espelhos devem ser posicionados de forma a serem visíveis pela* Kinect *e orientados de modo a refletirem a área de captura para o sensor.*

tem um alcance limitado e a utilização da reflexão entre espelhos aumenta a distância dos pontos capturados (nestas perspetivas) à câmara, o espaço disponível para realizar a captura também diminuiria. No entanto, dependendo do posicionamento dos espelhos, este tipo de reflexões pode continuar a existir. De forma a descartar estas reflexões e de modo a evitar mais restrições físicas na colocação dos espelhos no espaço, o mapa de profundidade produzido pela *Kinect* será filtrado pela distância. Assim, os pontos que se encontram a distâncias superiores, resultado desta dupla reflexão, serão excluídos.

Esta arquitetura genérica permite a criação de várias configurações para a realização da captura 360° de objetos. No entanto, foi necessário implementar casos específicos desta mesma arquitetura. Desta forma, foram estudadas algumas possibilidades, e foram implementadas e testadas duas abordagens diferentes a esta arquitetura. As fotografias das configurações construídas bem como dos momentos da captura podem ser encontradas neste link[1].

4.1.1 Caso 1

O primeiro protótipo construído utiliza a *Kinect* e dois espelhos para realizar a captura 360° criando desta forma três pontos de captura distintos. Nesta configuração os elementos estão dispostos a distâncias semelhantes da área de captura formando um triângulo centrado nessa área.

Para os testes realizados, a *Kinect* foi colocada a uma altura de cerca de 120cm, com uma inclinação aproximada de 35°, de forma a estar apontada para o centro daquilo que será a área de captura, e a uma distância de 170cm do mesmo. Por sua vez, os espelhos foram colocados do lado oposto à *Kinect*, de forma a ocuparem as margens laterais do campo de visão do sensor, pousados no chão e a uma distância do centro da área de captura de aproximadamente 100cm. A orientação destes materiais foi ajustada manualmente de modo a refletir o máximo de informação da área de captura em direção à câmara (Figura 4.4).

A dimensão da área de captura está relacionada não só com a distância a que os elementos se encontram entre si mas também com o tamanho das entidades a capturar. Neste caso de estudo, o centro desta área corresponde aproximadamente ao ponto central entre todos os elementos. A primeira restrição espacial deriva do limite de alcance mínimo da *Kinect*, o que significa que a captura apenas se realiza a partir de aproximadamente 50cm da posição do sensor. Além disso, e de forma a poder utilizar a informação dos espelhos, o objeto não pode ocludi-los completamente, como tal, a dimensão deste não pode ocupar todo o campo de visão do sensor. Já do lado dos espelhos,

[1] https://www.dropbox.com/sh/v8vjwvilbibchoq/AABdqm0RBpNWTaPyNEn5ZIZla?dl=0 (acedido em outubro de 2014)

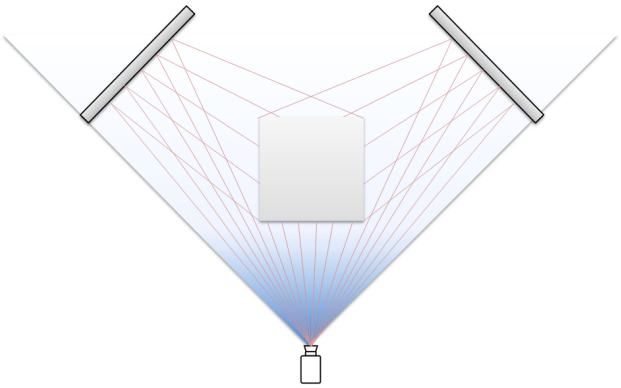


Figura 4.4: Exemplo das reflexões efetuadas pelos espelhos de forma a capturar a informação presente na área de captura.

essa restrição não existe, no entanto, caso o objeto se encontre muito próximo, o espelho pode não ter ângulo suficiente para refletir a informação da parte de trás do objeto para o sensor. Como tal, apesar de estar relacionado com a dimensão do objeto a capturar, revelou-se útil existir uma margem, com aproximadamente 40cm, de modo a poder utilizar corretamente a informação proveniente destes pontos de captura virtuais. De forma esquemática, uma aproximação da área de ação do presente caso de estudo pode ser vista na Figura 4.5. Uma vez que durante a captura se lida com informação em 3D, esta área de captura não pode ser vista apenas como uma delimitação em 2D mas sim como o volume estendido a partir dessa base.

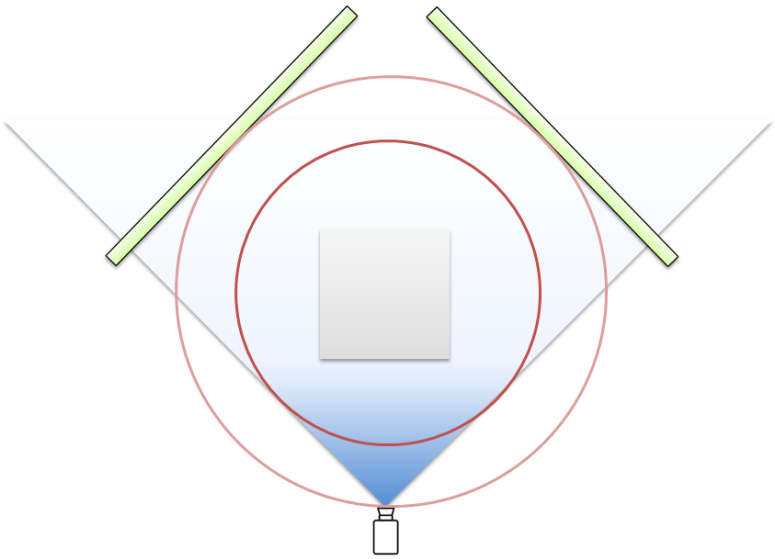


Figura 4.5: Esquema 2D em vista de topo da definição da área de ação.

O detalhe da informação capturada nesta configuração está dependente da perspetiva pelo qual esta é adquirida. No caso da informação capturada diretamente pelo sensor o erro é menor e está apenas relacionado com a distância da superfície à câmara. Assumindo que a superfície do objeto se encontra entre os 50cm e os 100cm, de acordo com a equação apresentada na Secção 3.3.2, o erro associado será de 0,17cm e de 0,76cm respetivamente. Já no caso da informação refletida pelos espelhos, uma vez que a distância à *Kinect* é maior, o erro é também maior. Assumindo que a superfície do objeto detetada pelo espelho também se encontra entre os 50cm e os 100cm deste, adicionando a distância desse mesmo espelho à *Kinect* (aproximadamente 250cm) obtemos uma distância absoluta de aproximadamente 300cm e 350cm. Nestes casos, e aplicando a mesma fórmula, o erro associado será de 6,9cm e 9,3cm respetivamente. Esta configuração é bastante versátil o que possibilita a criação de uma área de captura bastante ampla. Desta forma, existe uma maior liberdade em relação à dimensão das entidades a serem capturadas e dos movimentos que estas podem efetuar. No caso de as entidades terem dimensões superiores à área capturada, este sistema pode ser adaptado através da colocação do material de aquisição em posições mais distantes, aumentando assim a área de aquisição. Esta abordagem está no entanto limitada às características de alcance máximo da *Kinect* e tem como maior desvantagem a perda de qualidade na informação inerente à distância a que o objeto se encontra do sensor, principalmente nas perspetivas resultantes da reflecção dos espelhos.

4.1.2 Caso 2

O segundo protótipo construído utiliza a *Kinect* e quatro espelhos para realizar a captura 360° criando assim cinco pontos de captura distintos. Nesta configuração os elementos estão dispostos em forma de pirâmide com o centro da área de ação a coincidir com o centro da base da mesma.

Neste caso a *Kinect* foi colocada no topo a uma altura de aproximadamente 160cm e apontada para o centro da área de captura, ou seja, numa perspetiva perpendicular ao chão. Os espelhos são colocados no chão de acordo com o campo de visão da *Kinect* e de forma a ocupar todas as margens do mesmo, tomando assim o papel de arestas na base da pirâmide. A orientação destes foi também ajustada manualmente criando um ângulo de aproximadamente 55°, de forma a refletir a informação da área de captura em direção à *Kinect* (Figura 4.6). Dada a disposição dos elementos, existe a sobreposição de informação entre espelhos, no entanto, e como já foi referido, esta pode ser filtrada posteriormente a partir do mapa de profundidade.

A área de captura deste caso de estudo está fisicamente limitada pelo posicionamento dos espelhos uma vez que estes demarcam uma área fechada em

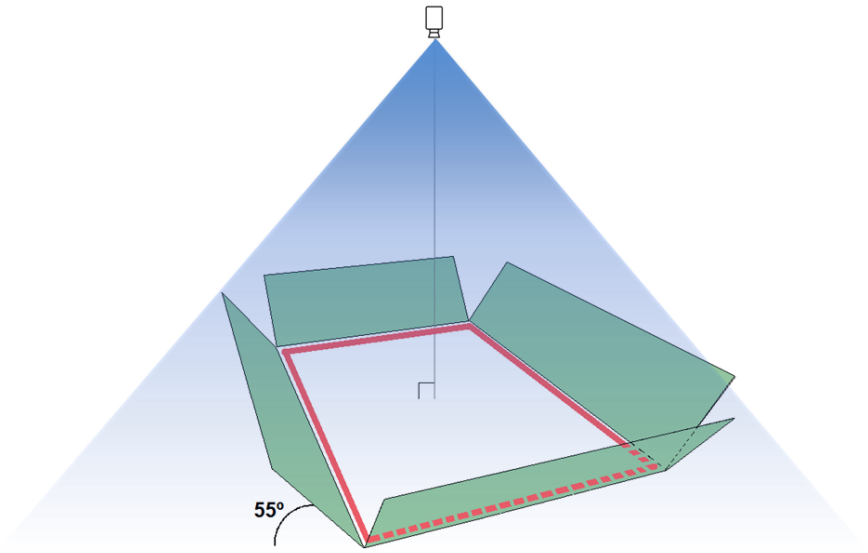


Figura 4.6: Esquema do posicionamento dos objetos na configuração do Caso 2.

forma de retângulo. No entanto, este não é o único fator limitativo e o volume dos objetos pode também restringir esta área. O seu posicionamento em zonas próximas aos espelhos resulta na oclusão dos mesmos e, como tal, limita a captura de informação. Desta forma, revelou-se útil a existência de uma margem à volta da área de captura com aproximadamente 10cm. Além disso, esta configuração tem também limitações a nível da altura dos objetos: se a superfície mais próxima do sensor estiver a uma distância inferior a 40cm da *Kinect*, esta já não a conseguirá capturar corretamente. Desta forma, o volume da área de captura desta configuração pode ser vista nos esquemas da Figura 4.7.

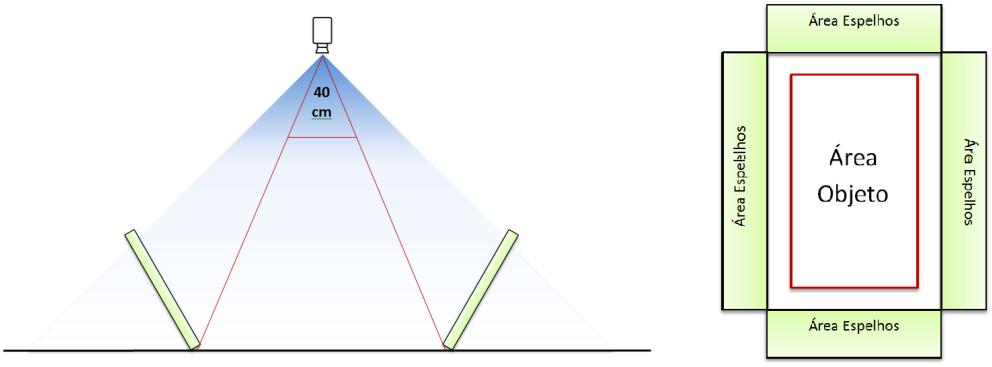


Figura 4.7: Esquema do campo de visão da Kinect *e das implicações relativamente à limitação da área de captura.*

O detalhe da informação capturada está igualmente dependente da perspetiva pela qual esta é adquirida e da distância a que os objetos se encontram do sensor. Se a captura for efetuada diretamente pela câmara, o erro associado a uma leitura depende apenas da distância a que a superfície se encontra da câmara. Aplicando, do mesmo modo, a fórmula descrita na 3.3.2, se esta se encontrar a 50cm da câmara o erro será de aproximadamente 0,17cm. Por outro lado, se a superfície do objeto estiver rente à base e, como tal, se encontrar a 150cm da câmara, o erro associado será de aproximadamente 17,3cm. Já no caso da informação refletida pelos espelhos, o erro associado a essas medições está dependente da distância do objeto ao espelho e da distância do espelho à *Kinect*. No caso dos espelhos que ocupam a base maior, a distância destes à câmara é de aproximadamente 15cm, enquanto que os espelhos da base menor se encontram a 35cm. Desta forma, um objeto cuja superfície a medir se encontre no centro da zona de captura, se a captura for efetuada por um espelho pertencente à base maior, esta encontra-se a 175cm da *Kinect* $(160 + 15)$ e, como tal, tem um erro associado de 2,35cm. Por outro lado, se a medição for feita a partir de um espelho que se encontra na base menor da área de captura, a superfície encontra-se a 200cm da câmara $(165+35)$ tendo um erro associado de 3cm.

Esta configuração limita um pouco mais a área de captura uma vez que o maior número de elementos e o posicionamento que estes tomam restringem a utilização do espaço. Por outro lado, a adição de um maior número de espelhos faz com que existam mais perspetivas a capturar e, como tal, é gerada mais informação distinta em simultâneo. Desta forma, esta configuração está mais orientada para a aquisição de objetos ou entidades de pequenas ou médias dimensões. Caso se pretenda utilizar o sistema para objetos maiores, é possível adaptá-lo através do reposicionamento da câmara a uma altura superior e ajustando os espelhos a essa nova posição. Tal como o primeiro caso de estudo, esta configuração está limitada ao alcance mínimo e máximo da *Kinect* pelo que as distâncias dos objetos ao sensor (direta ou indiretamente) devem ser superiores a 0,5m e inferiores a 3,5m. Além disso, a qualidade da informação capturada pela *Kinect* é igualmente influenciada pela distância a que os objetos se encontram do sensor uma vez que, quanto maior a distância, maior o erro associado.

4.2 Fluxo de execução

Independentemente da abordagem/configuração escolhida, o fluxo de execução deste sistema segue os mesmos princípios base, sendo que a principal variação é o número de espelhos disponíveis para simular pontos de aquisição virtuais.

A nível lógico, o primeiro passo deste fluxo é um passo único e consiste na

calibração do sistema, isto é, no posicionamento da *Kinect* e dos espelhos e no estabelecimento das relações entre eles. Neste passo é criada a informação que permite a correta geração de informação 3D durante a captura. Depois de concluída a calibração, o sistema está pronto a iniciar o ciclo de captura de informação. O segundo passo, primeiro deste ciclo, consiste na captura de informação. Nele, além de capturada, a informação é ainda filtrada de modo a separar aquela que é gerada diretamente pela câmara da informação proveniente dos espelhos. Esta última terá que ser tratada de forma diferente para, juntamente com a informação de calibração, gerar o correto posicionamento dos pontos no mundo 3D. Esta informação é depois processada, no terceiro passo, de forma a gerar o resultado final, o que pode passar pela remoção de ruído, simplificação da nuvem de pontos ou até a geração de uma malha poligonal. No fim deste processo e depois de criada a informação das entidades em 3D, resta visualizá-la, o que constitui o quarto passo. Além ser possível observar o resultado final da captura em tempo real, é neste passo que se pode validar os resultados obtidos.

A nível de concretização, de forma a tornar o sistema mais eficiente, foi desenvolvida uma arquitetura com três fluxos de execução paralelos (*threads*), um para cada fase do ciclo de captura (Figura 4.8). O primeiro é responsável apenas por ler a informação proveniente da *Kinect* e disponibilizá-la para os passos seguintes, ficando assim livre de qualquer atraso provocado por processamento de informação. O segundo fluxo (o fluxo principal), depois de ter uma cópia da informação da *Kinect*, aplica todo o processamento necessário de forma a construir a nuvem de pontos final do objeto. Este é um passo computacionalmente exigente e como tal, lento. O terceiro fluxo de execução é responsável pela visualização da informação e permite que esta seja feita sem qualquer degradação de *performance* resultante do processamento de informação, o que permite uma experiência de visualização e navegação mais fluida.

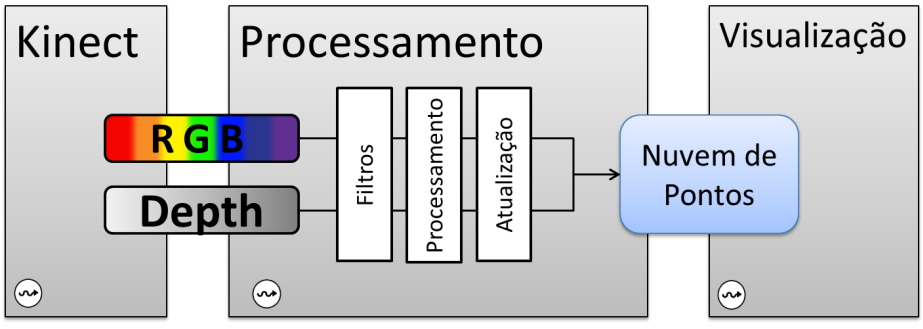


Figura 4.8: Esquema do fluxo de execução implementado.

De forma a detalhar o processo seguido por este sistema, o fluxo de exe-

cução será dividido nas quatro fases lógicas, que serão descritas nas próximas subsecções. Para facilitar a perceção das explicações, os exemplos dados serão referentes ao Caso 1, constituído por uma *Kinect* e dois espelhos.

4.2.1 Configuração

O primeiro passo consiste na configuração e calibração do sistema. Este é realizado apenas uma vez e nele é gerada a informação necessária para, nas fases seguintes, gerar a informação tridimensional e possibilitar a junção da informação das diferentes perspetivas.

Depois de posicionar o material de acordo com o descrito no Caso 1, para o sistema funcionar é necessário definir quais as áreas que correspondem à área de captura e quais as áreas ocupadas pelos espelhos. Esta definição passa por, na imagem obtida pela *Kinect*, delimitar a área de ação de cada um deles e calcular o plano que os representa. Nos dois casos o processo de delimitação da área é feito da mesma forma. É gerada uma imagem que contém a informação *RGB* da cena e a máscara do mapa de profundidade e nela são selecionados os pontos que delimitarão a área pretendida. No caso dos espelhos corresponde às áreas da imagem onde o espelho é visível e onde aparecerá a reflecção dos objetos a capturar. Já no caso da área de aquisição, esta delimitação corresponde ao plano onde os objetos ou entidades serão pousadas para serem analisadas pelo sistema, e a toda a área da imagem que o volume da zona de aquisição pode ocupar (Figura 4.9). Uma vez que esta área é mais extensa, é comum que esta se sobreponha com as áreas delimitadas pelos espelhos. No entanto, esta sobreposição não é problemática uma vez que a informação será filtrada através da distância a que os pontos se encontram da *Kinect*.

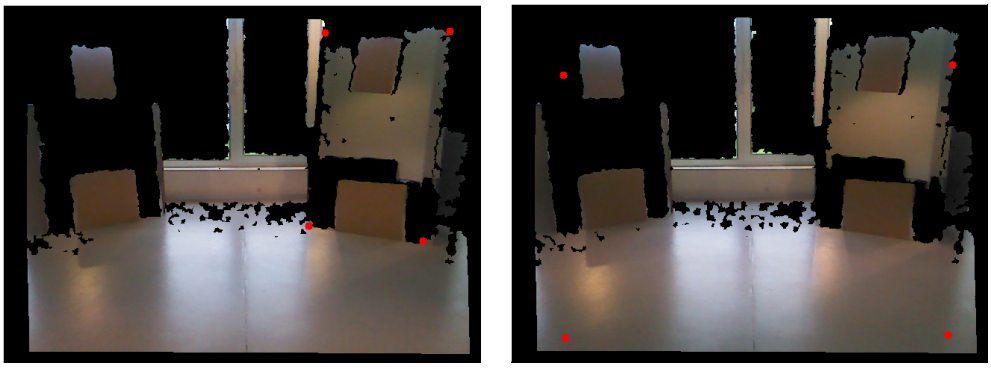


Figura 4.9: Seleção da área correspondente a um dos espelhos (esquerda) e da área de captura (direita). Esta última é mais extensa e ocupa quase toda a imagem uma vez que corresponde à visão frontal de um volume e não só à delimitação de um plano.

A extração do plano, no caso da área de captura, pode ser feita de duas formas distintas, uma automática ou a outra, manual. A extração automática é efetuada a partir do *OpenNI* através de uma funcionalidade que permite calcular o plano do chão segundo alguns pressupostos como a posição da câmara (horizontal em relação ao piso) e a área de chão visível. No entanto nem sempre é possível realizar esta extração e, como tal, poderá ser necessário calculá-lo de forma manual. Esta extração é feita através da seleção de pontos na imagem pertencentes ao plano que se pretende calcular. Depois de selecionados, estes são convertidos para um referencial 3D, gerando assim pontos tridimensionais, e a partir destes e utilizando o algoritmo de *Ransac* (*RANdom Sample Consensus*)[Fischler and Bolles, 1981] é calculado o plano que melhor se adapta à amostra. Este plano será usado como o plano representativo da área de captura.

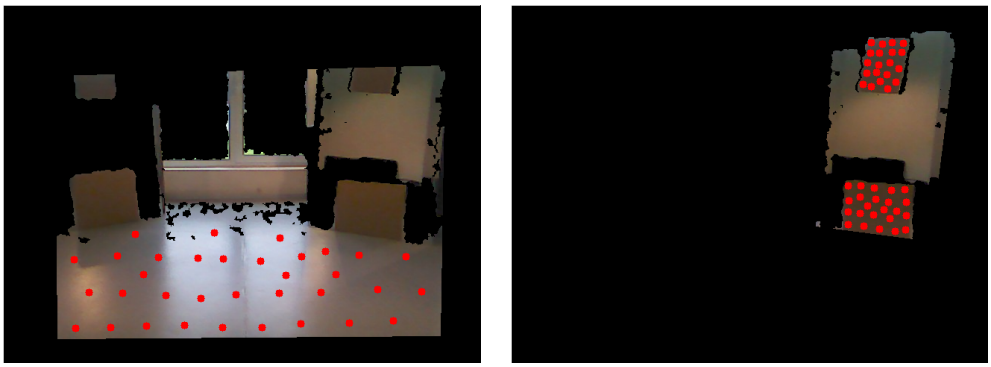


Figura 4.10: À esquerda, seleção dos pontos pertencentes ao chão para cálculo manual da equação do seu plano. Na direita, seleção dos pontos pertencentes aos artefactos colocados no espelho para efetuar o cálculo da equação do plano desse espelho.

No caso dos planos correspondentes aos espelhos esta seleção não pode ser feita diretamente. Uma vez que os pontos presentes na imagem na área do espelho não contêm informação de profundidade do mesmo, apenas da sua refleção, a extração direta da posição 3D do plano do espelho torna-se impossível. De forma a contornar este problema foi necessário introduzir um artefacto, apenas presente nesta fase de calibração, que consiste num ou mais planos opacos (folhas de papel ou cartolina) colados ao espelho (Figura 4.10). Depois da colocação desse artefacto, a extração do plano é feita da mesma forma que a extração manual da área de captura e, neste caso, apenas é usada a área ocupada pelo(s) artefacto(s). Nele são selecionados e convertidos os pontos para o mundo 3D e posteriormente extraído o plano que representará toda a área ocupada pelo espelho.

Os planos calculados serão depois utilizados a processar a informação cap-

turada e gerar a correta posição dos pontos 3D como será explicado nas próximas secções. No fim deste passo, a fase de calibração está concluída e pode ser iniciado o ciclo de captura de informação.

4.2.2 Aquisição e pré-processamento

Depois da configuração, o primeiro passo do ciclo de captura de informação é a aquisição da informação. Isto passa pela obtenção do mapa de profundidade produzido pela *Kinect* e pela sua transformação numa nuvem de pontos tridimensional, recorrendo à informação de calibração intrínseca do sensor, presente no *OpenNI*[ROS.org, 2012]. No entanto, a informação capturada contém ruído que pode ser eliminado ainda antes da geração da nuvem de pontos de forma a melhorar os resultados e a *performance* da aplicação. Nesta fase de emagrecimento da nuvem de pontos é essencial a informação recolhida na fase anterior, a de configuração.

O primeiro passo é o acesso ao mapa de profundidade da *Kinect* que, como foi referido anteriormente, contém a informação da distância a que os elementos da cena se encontram da câmara. No entanto, esta informação encontra-se numa perspetiva de projeção e como tal, de forma a gerar a informação em 3D numa perspetiva ortogonal, a informação do mapa de profundidade tem que ser convertida. Este processo é feito de forma transparente para o utilizador relativamente aos parâmetros intrínsecos e extrínsecos da *Kinect* tendo sido utilizadas as configurações *standard* do *OpenNI*. Estes parâmetros podem ser recalibrados se necessário, no entanto nos testes efetuados tal não foi necessário e foram usadas as informações pré-definidas.

Uma vez que nem toda a informação do mapa será utilizada, esta pode ser filtrada ainda no mapa de profundidade, isto é, antes de ser convertido para 3D. A vantagem de realizar este processamento em 2D prende-se pelo facto de a velocidade de execução ser superior comparativamente ao processamento em 3D.

O primeiro filtro a ser aplicado corresponde à máscara construída a partir das áreas definidas na fase de calibração, tanto para os espelhos como para a área de captura. Esta máscara permite eliminar alguma informação irrelevante para o sistema, normalmente situada na periferia da imagem e nas zonas de junção entre elementos (espelhos e chão). O segundo filtro a ser aplicado nesta fase está relacionado com a distância a que os objetos se encontram da câmara nas diferentes áreas. No caso retratado em Caso 1, a área de ação situa-se entre os 150cm e 250cm de distância da *Kinect*, como tal, a informação que se encontre fora destes limites pode ser descartada. No caso dos espelhos, além de diminuir a quantidade de informação a ser processada, esta filtragem é desejável também para descartar os objetos que se encontrem entre a câmara

e os espelhos, de forma a garantir o correto processamento dos mesmos. Estes casos serão tratados como informação vista diretamente pela câmara e não como um reflexo do espelho. Ambos os casos são ilustrados na Figura 4.11.

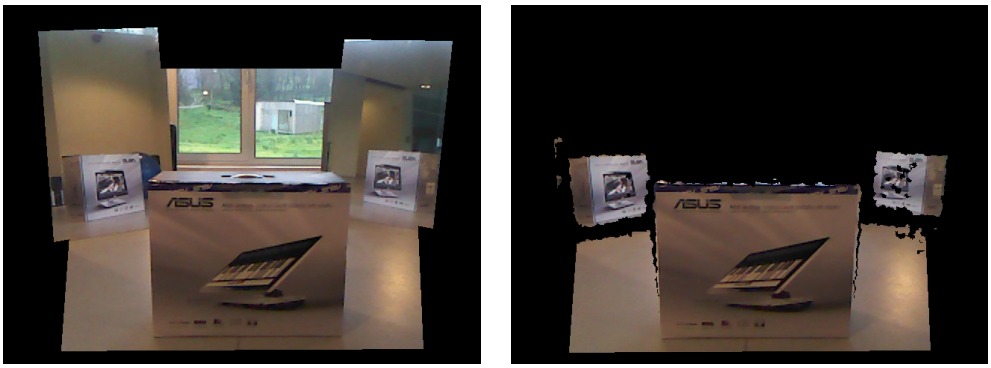


Figura 4.11: Exemplos de máscaras simples aplicadas à imagem. À esquerda, máscaras definidas na fase de calibração. À direita, máscara criada com o filtro de distância.

Além destes filtros, é possível aplicar algum pré-processamento ao mapa de profundidade antes de transformar a informação que este contém uma nuvem de pontos. Este pré-processamento passa essencialmente pela aplicação de métodos de processamento de imagem ao mapa de profundidade de forma a suavizá-lo e, como tal, tentar corrigir ou minimizar as imperfeições inerentes à captura da *Kinect*. No entanto, a aplicação destes métodos é computacionalmente exigente o que pode fazer com que o sistema perca a propriedade de tempo real. Estes métodos serão analisados em mais detalhe na secção seguinte.

Depois de finalizado este processo e de a informação ter sido filtrada, segue-se a transformação da informação para 3D com perspetiva real. Nos pontos em que existe visualização direta pela *Kinect*, a transformação destes para o mundo 3D é feita diretamente pelo *OpenNI*. Por outro lado, se os pontos correspondem a informação resultante da reflexão de um espelho, depois de convertidos estes têm que sofrer outra transformação de forma a ficarem corretamente posicionados na cena (Figura 4.11). De acordo com a fórmula descrita na Secção 3.3.1, é calculada a distância do ponto ao plano do espelho que o refletiu e depois, juntamente com a normal desse plano é calculada a posição correspondente do ponto. Apesar de ainda conter muito ruído, no fim destes dois passos obtém-se a primeira versão na nuvem de pontos da cena.

No caso de captura de um modelo estático a partir de múltiplas *frames* existe ainda outra fase no pré-processamento realizado antes da conversão da informação para 3D, que passa pela criação de um só mapa de profundidade resultante de todos os mapas capturados. Este passo tem duas vantagens ime-

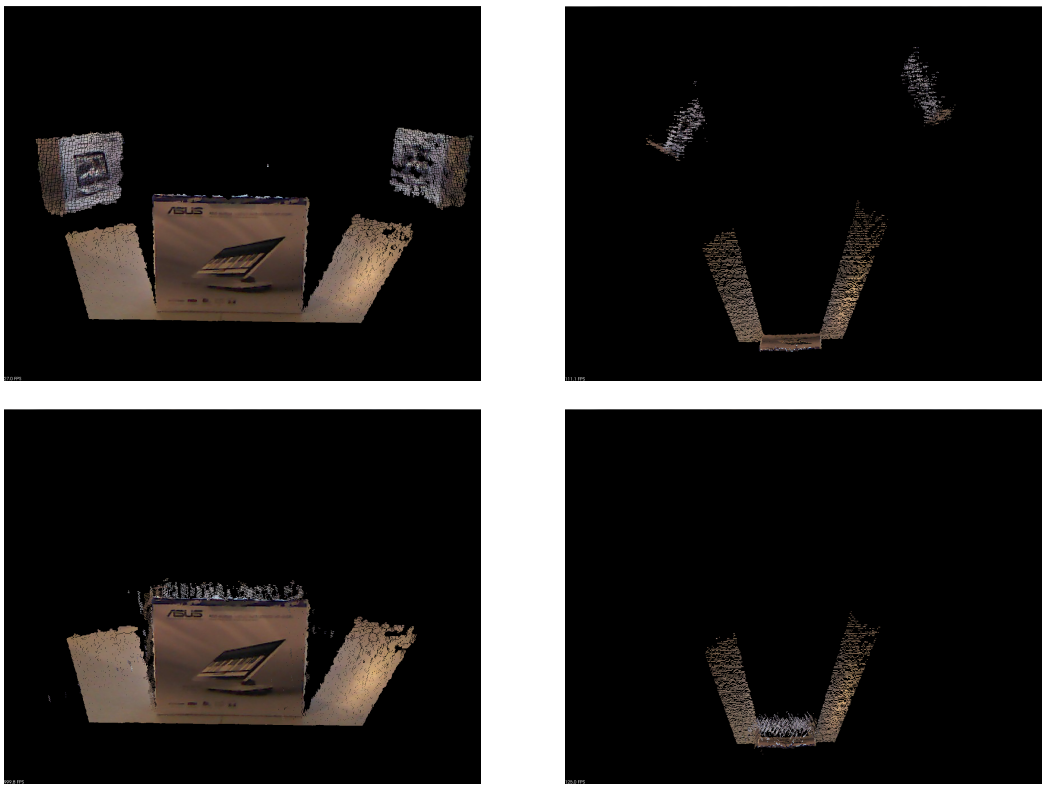


Figura 4.12: Exemplo da imagem 3D da captura com e sem reflexão dos espelhos. No topo a perspetiva frontal e vista de cima sem reflexão dos espelhos. Em baixo, a perspetiva frontal e vista de cima com reflexão dos espelhos.

diatas que correspondem ao preenchimento de "buracos" de informação e à suavização da mesma. A primeira acontece uma vez que a *Kinect* introduz esporadicamente falhas de informação no mapa de profundidade no entanto estas são geralmente corrigidas na *frame* seguinte. Desta forma e com a aquisição de múltiplas *frames*, a criação de um mapa de profundidade único permite que estas falhas sejam colmatadas. Já a suavização da informação passa pelo cálculo da média, ponto a ponto, de todos os mapas de profundidade, o que permite estabilizar possíveis imprecisões na aquisição do mapa de profundidade. Tanto num caso como no outro, estas melhorias apenas se podem efetuar em cenas estáticas uma vez que qualquer movimento causaria alterações significativas no mapa de profundidade e, como tal, resultados errados na nuvem de pontos 3D final.

4.2.3 Processamento de informação

O segundo passo do ciclo da captura de informação é o processamento da mesma, já em 3D. O objetivo desta fase é a eliminação de ruído, isto é, das partes do cenário que não pertencem ao objeto que se quer capturar, e a criação e suavização de uma malha de pontos representativa do mesmo objeto. Observando-se o resultado em 3D da fase de aquisição (Figura 4.12), existe ainda muita informação desnecessária nomeadamente da superfície onde o objeto se encontra pousado. Esta informação pode ser descartada de forma fácil uma vez que durante a fase de calibração foi recolhida informação sobre o plano do chão. Desta forma, foi estabelecida uma margem (3,5cm neste caso) e todos os pontos que se encontrem a uma distância inferior a essa margem são considerados pontos do chão e como tal, eliminados (Figura 4.13).

Figura 4.13: Perspetiva frontal (esquerda) e vista de cima (direita) da cena com a remoção do chão.

Depois deste passo, a informação disponível pertence quase toda aos objetos encontrados na área de captura, no entanto ainda é possível encontrar algum ruído. Isto deve-se a imprecisões nas medições efetuadas pela *Kinect* e/ou aos valores adquiridos durante a fase de calibração que podem também eles não ser exatos. De forma a remover este ruído de modo autónomo é necessário aplicar algoritmos diretamente à nuvem de pontos, tanto para remover *outliers*, isto é, pontos que não pertencem ao objeto, como para uniformizá-los, suavizando a superfície do mesmo. No entanto, estes passos acarretam uma carga computacional elevada e, como tal, foi impossível ao sistema aplicá-los e manter a propriedade de processamento em tempo real. Já no caso da captura de informação estática, estes métodos podem ser aplicados e posteriormente o ruído pode ainda ser removido manualmente com auxílio a outras ferramentas como o *MeshLab*. Estas soluções serão abordadas na secção seguinte em mais detalhe.

Tendo a nuvem de pontos gerada, outro passo que pode ainda ser efetuado é a criação de uma malha poligonal que represente a superfície do objeto da

captura. Apesar de este não ser um dos objetivos do sistema, foi implementado um método para realizar esta tarefa utilizando o *PCL* e um algoritmo de triangulação "gananciosa" [Marton et al., 2009]. No entanto, e assim como os métodos de processamento em 3D, a geração da malha poligonal a partir da nuvem de pontos é um processo exigente e, como tal, impossível de realizar em tempo real.

4.2.4 Visualização de informação

O último passo do ciclo consiste na visualização da informação gerada. Esta pode ser feita em tempo real recorrendo às bibliotecas disponibilizadas pelo *PCL*, ou através da visualização de modelos estáticos com ferramentas como o *MeshLab* ou o *Blender*.

A visualização da informação em tempo real possibilita a perceção imediata daquilo que o sistema consegue capturar e permite ter uma primeira ideia do que poderá ser o resultado final. Além da simples navegação no cenário e visualização do mesmo a partir de diferentes pontos de vista, esta vertente é também útil enquanto ferramenta de configuração para a captura de objetos. Isto permite observar o ajuste de alguns parâmetros de configuração, como o limite de distâncias ou até mesmo as equações dos planos dos espelhos, e permite analisar se existem sobreposições entre objetos, possibilitando o reposicionamento dos mesmos para uma captura com condições mais favoráveis. Outra vantagem desta visualização prende-se pela possibilidade de, em tempo real, alterar e observar o efeito de filtros ou algoritmos de processamento sobre a informação capturada de forma a otimizar a qualidade dos resultados.

Outra das funcionalidades do sistema é a geração de nuvens de pontos 3D representativas dos objetos e como tal é necessário passá-las para um formato persistente. Desta forma, uma das opções disponibilizada pelo sistema desenvolvido é a possibilidade de gravar essas nuvens para ficheiro, sendo possível escolher entre os formatos *ply*, *obj* ou então, o formato próprio do *PCL*, o *pcd*. Os exemplos disponibilizados foram guardados no formato *ply* por ser um formato genericamente bem aceite e uma vez que o formato *obj* é mais orientado a *meshes* e não a nuvens de pontos.

A visualização deste tipo de ficheiros pode ser feita em várias aplicações como o *MeshLab* ou o *Blender*. Além da sua visualização, outro objetivo da criação destes modelos é a sua validação sendo para isso necessária a análise da qualidade dos mesmos. Isto passa por, em comparação com o objeto real, verificar as medidas dos modelos, analisar a existência de buracos e as diferenças entre as superfícies e a presença de *outliers*, isto é, informação que não pertence ao objeto. Para realizar estas tarefas, e também algumas de correção (como a remoção desses mesmos *outliers*), foi escolhido o *MeshLab*. As ra-

zões dessa escolha assim como as suas características estão descritas em mais detalhe na secção de Tecnologia.

4.3 Problemas e soluções

Tal como já foi sendo descrito na secção anterior, durante a implementação do sistema surgiram várias questões, tanto na fase de aquisição, como na geração da nuvem de pontos tridimensional. Alguns destes problemas puderam ser atenuados, ou até resolvidos, através de técnicas simples, como a utilização de filtros por máscara e por distância, no entanto, o resultado final continua a conter várias imperfeições.

De forma a analisá-las, estas imperfeições foram classificadas em três tipos: informação imprecisa, ruído e falhas de informação. Além de as detalhar, nas próximas secções serão explicadas as origens destas imprecisões e analisadas possíveis soluções ou formas de melhorar a qualidade da informação.

4.3.1 Informação imprecisa

Uma das dificuldades encontradas é inerente ao *hardware* utilizado. Apesar de conseguir capturar informação de profundidade com detalhes na ordem dos milímetros, a *Kinect* guarda essa informação de forma discreta e não linear, como foi mostrado no Gráfico 3.6 da Secção 3.3.2, o que faz com que o erro aumente exponencialmente com a distância. Esta característica, juntamente com outros fatores como a iluminação ou as propriedades das superfícies a capturar, faz com que o mapa de profundidade produzido tenha uma qualidade inferior à desejada e provoque resultados facetados como os representados na Figura 4.14.

A diferença no posicionamento dos pontos deve-se a esta discretização do mapa de profundidades. Utilizando um caso concreto, imagine-se uma superfície plana perpendicular à orientação da câmara. Neste caso, depois de convertidos para 3D, todos os pontos deviam encontrar-se no mesmo plano, no entanto, uma vez que os pontos da superfície não se encontram à mesma distância do sensor, isto não acontece. Se os pontos se encontrarem entre 1994 e os 2005 milímetros de distância, o valor no mapa será de 916 enquanto os pontos que se encontrem entre os 2006 e os 2017 milímetros, o valor será de 917. Depois de convertidos para 3D, os pontos do primeiro intervalo serão todos posicionados a uma distância de 2000 milímetros e os do segundo a 2011,5 milímetros, provocando as descontinuidades e o aspeto facetado visível na Figura 4.14. Estas imprecisões agravam-se com a distância a que os objetos se encontram da *Kinect*, o que faz com que, por exemplo, a informação recolhida

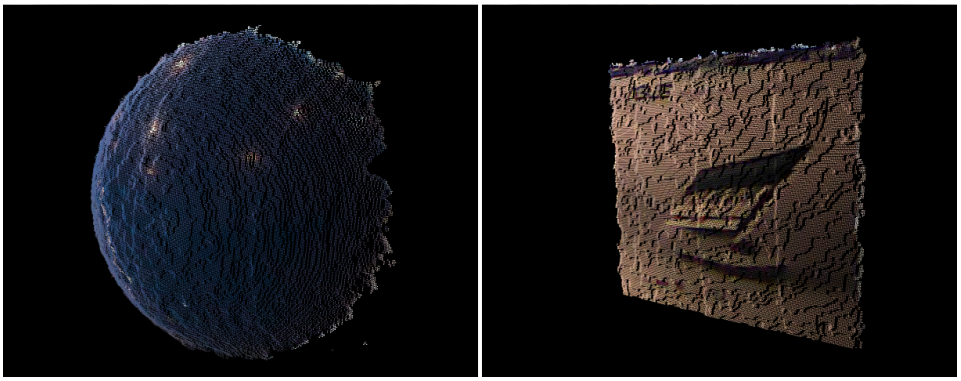


Figura 4.14: Ilustração das imprecisões inerentes às limitações da Kinect. *O aspeto facetado das superfícies deve-se à forma linear como a informação é guardada pela* Kinect.

pelos espelhos tenha menos qualidade que a informação recolhida diretamente pelo sensor.

Uma vez que estes erros acontecem na base de todo o processo da captura e, como tal, são propagados para os restantes passos, é importante que sejam tratados também numa fase inicial. Apesar de não ser possível corrigi-los totalmente, é possível minimizar este problema através da aplicação de filtros de suavização ao mapa de profundidade. Genericamente, estes filtros utilizam um método que consiste na análise dos píxeis vizinhos de cada píxel e no cálculo de um valor representativo para essa amostra. O número de píxeis vizinhos a considerar é parametrizável, assim como a métrica utilizada para obter esse valor representativo. De forma a minimizar este problema, os filtros testados foram:

- Filtro Médio - todos os valores dos píxeis vizinhos têm a mesma importância e, como tal, o valor representativo é uma média de todos eles. Uma vez que a métrica usada é simples, este é o filtro com melhor desempenho.

- Filtro Gaussiano - os valores dos píxeis têm pesos diferentes para o valor final e esse peso segue uma distribuição gaussiana, isto é, os píxeis centrais (vizinhos mais próximos) têm uma importância maior que os píxeis da periferia (vizinhos mais distantes). Apesar do seu comportamento ser linear, este filtro requere mais cálculos que o anterior e, como tal, a sua *performance* é ligeiramente inferior.

- Filtro Bilateral - enquanto que os dois filtros anteriores se comportam sempre da mesma forma independentemente da imagem, este filtro, além de considerar a distância dos píxeis vizinhos, também considera a sua

intensidade para calcular a sua importância. A análise das diferenças de intensidade entre píxeis vizinhos permite identificar a existência de arestas nas imagens e salvaguardar essas propriedades em vez de suavizalas [Tomasi and Manduchi, 1998]. Uma vez que o cálculo necessário para este filtro é mais complexo, o tempo de execução é também pior que o dos anteriores.

Foram realizados testes com os três filtros (Figura 4.15) e o filtro que obteve melhores resultados foi o filtro Bilateral. O filtro Gaussiano foi descartado uma vez que introduz ruído no mapa de profundidade reduzindo a qualidade do mesmo. O filtro Bilateral teve melhores resultados que o filtro Mediano apresentando resultados mais suaves e menos deformações na zonas correspondentes às extremidades dos objetos.

A nível de desempenho e apesar da diferença de métodos, todos os filtros obtiveram resultados semelhantes, diminuindo a taxa de atualização em aproximadamente 40%. Tendo em conta a diferença de complexidade entre os filtros Mediano e Bilateral, a diferença de *performance* entre ambos foi comparada e foi verificada uma maior diminuição da taxa de atualização no caso do filtro Bilateral (cerca de 5%).

4.3.2 Ruído

Enquanto que a informação imprecisa se identifica essencialmente pelas imperfeições nas superfícies dos objetos capturados, o ruído consiste na informação da nuvem de pontos que não faz parte do objeto mas que não foi possível filtrar. Este ruído pode ter diversas fontes e pode ser introduzido em qualquer uma das fases.

Na fase inicial de captura, isto é, no momento em que a informação é capturada pela *Kinect*, pode haver a introdução de ruído como consequência de condições menos favoráveis à aquisição. Isso pode resultar da existência de materiais especulares ou transparentes/semitransparentes, existência de luzes fortes, etc. Estas condições são adversas à captura e podem fazer com que haja a introdução de valores errados nesta fase. Embora muitas vezes estas condições se traduzam em buracos de informação (Secção 4.3.3), por vezes, pode haver a introdução de valores válidos, mas incorretos. Isto faz com que haja a perceção no mapa de profundidade que partes de um determinado objeto se encontrem a uma distância diferente do que realmente está, provocando desta forma ruído na informação (Figura 4.16 (1)). Já na fase de processamento de informação, a introdução de ruído acontece quando, depois da aplicação dos filtros, existe ainda informação que devia ter sido filtrada. Além da possível existência de informação incorreta proveniente da captura, outro motivo para os filtros não conseguirem reter toda a informação desnecessária passa por

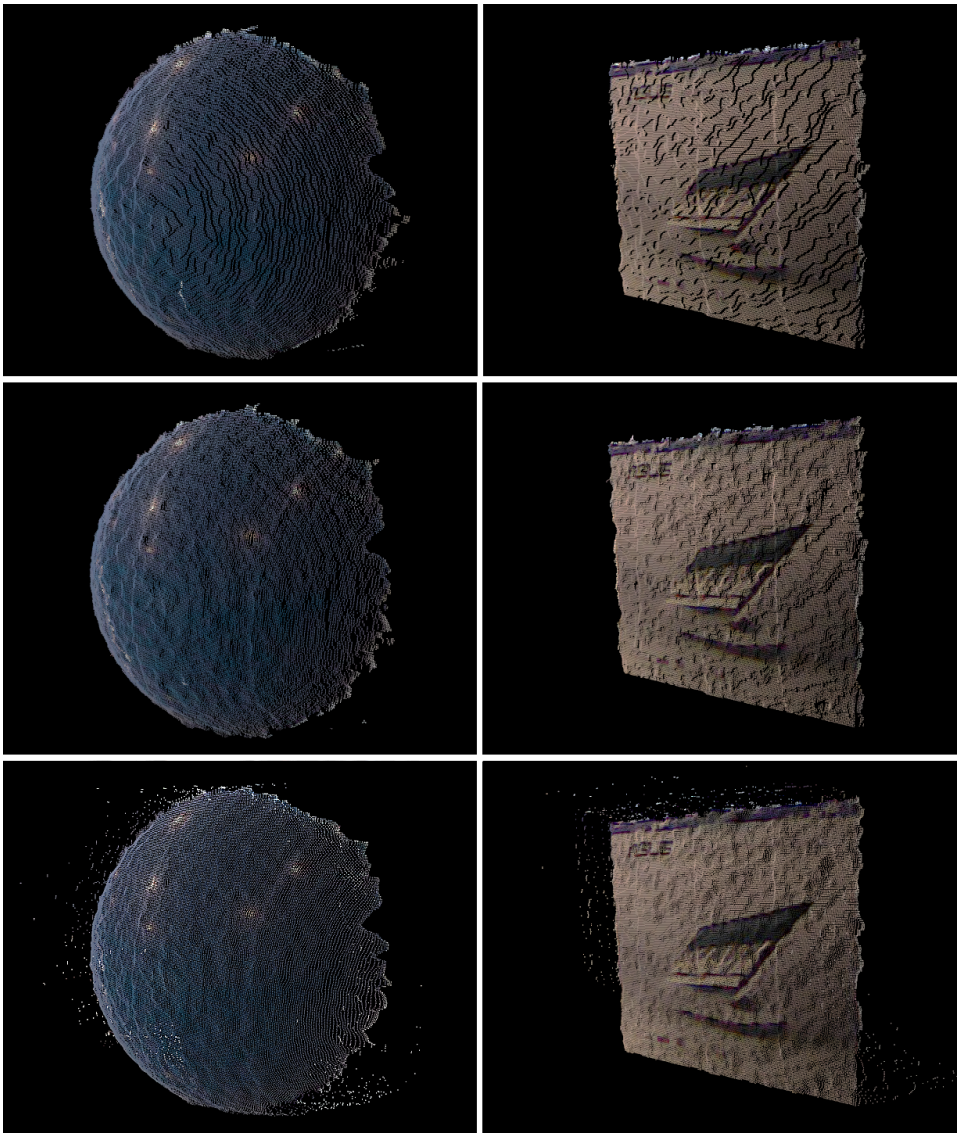


Figura 4.15: Ilustração da aplicação do filtro Mediano (topo) filtro Bilateral (centro) e o filtro Gaussiano (baixo).

imprecisões na fase de calibração. Isto tem como resultado direto a colocação incorreta dos pontos provenientes das reflexões dos espelhos ou, no caso do chão, uma criação de um plano do mesmo errado. Acontecendo isto, quando os filtros são aplicados não são totalmente eficazes e deixam passar informação que, caso a calibração tivesse sido correta, teria sido filtrada (Figura 4.16 (2)). Estes pontos são comummente chamados de *outliers*.

Existe ainda outro tipo de ruído que se apresenta de forma diferente dos *outliers* e que consiste na presença de informação em excesso numa determinada área. Na fase de aquisição de informação e geração da nuvem de pontos, além

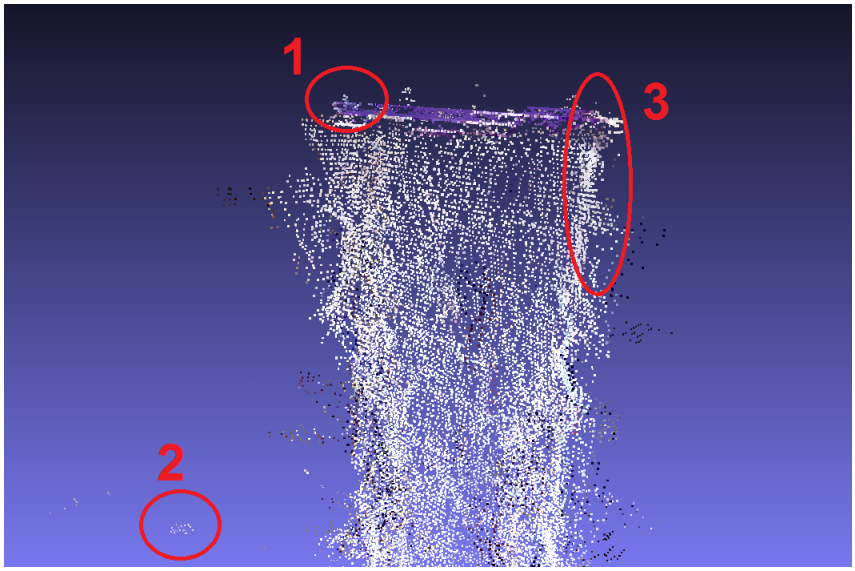


Figura 4.16: Demonstração dos diferentes tipos de Ruído durante a captura da Caixa da Kinect. *Em (1), ruído na profundidade da mesh, (2), existência de* outliers, *(3), sobreposição de informação das várias capturas.*

da calibração inicial, por questões de desempenho, não é efetuado qualquer outro cálculo para juntar a informação das diferentes perspetivas. Assumindo uma calibração correta, os pontos são, também eles, corretamente mapeados para a sua posição 3D, formando assim a nuvem de pontos da cena. Quando duas ou mais perspetivas conseguem adquirir informação de uma determinada porção de superfície, todas elas contribuirão para a construção da nuvem de pontos dessa mesma zona. Comparativamente com outras áreas, este fenómeno pode gerar um número elevado de pontos nessa área o que, no caso da criação da *mesh* e na análise da superfície, pode provocar mais ruído. Isto pode ser causado tanto pelas diferentes distâncias entre fontes de captura (e consequentemente, diferentes níveis de detalhe da captura) como por pequenas incorreções da calibração (Figura 4.16 (3)).

Nos dois primeiros casos a remoção do ruído passa por tentar eliminar estes *outliers*. Normalmente estes são pequenas porções de informação separadas ou ligadas apenas por uma pequena área ao objeto principal. O objetivo é identificá-las e eliminá-las. Existem três formas de realizar este processo, 2D, 3D e Manual:

- A remoção em 2D consiste no processamento do mapa de profundidade, já filtrado, com recurso ao *OpenCV*. Utilizando o método para encontrar contornos é possível listar todas as áreas do mapa em que existe informação útil. Depois, analisando o número de pontos pertencentes a essas áreas é possível descartar as mais pequenas que, tipicamente,

correspondem aos *outliers*.

- A remoção em 3D passa pelo processamento da nuvem de pontos e, com o auxílio aos métodos já existentes para este fim disponibilizados pela biblioteca *PCL*, encontrar e eliminar os *outliers*. Existem dois métodos disponíveis, um através da contagem dos pontos mais próximos e outro através de uma análise estatística dos mesmos. Ambos são parametrizáveis, permitindo definir métricas como o raio de procura ou o desvio padrão, e têm como resultado final uma nuvem de pontos filtrada, isto é, sem a maior parte dos *outliers*.

- A remoção manual consiste na seleção e eliminação destes pontos. Este processo apenas pode ser realizado numa fase final, isto é, em modelos cuja aquisição já tenha sido finalizada, e através da utilização de ferramentas como o *MeshLab* ou o *Blender*. Estas permitem a seleção de pontos (individualmente ou em conjunto) e eliminá-los da nuvem de pontos de forma a aperfeiçoar o resultado final.

As duas primeiras soluções, apesar de parametrizáveis, são soluções automáticas e, como tal, podem não ser totalmente eficazes não conseguindo remover todo o ruído. A primeira solução, apesar de mais rápida que a segunda a nível de execução (é processada em 2D), além de remover ruído pode também remover, erradamente, informação correta sobre o objeto e, como tal, foi descartada. Como exemplo, um caso onde isto pode acontecer é na informação proveniente dos espelhos, uma vez que, devido à existência de outros ruídos, esta pode ser reduzida e, como tal, confundida com ruído e eliminada. A segunda solução é mais custosa a nível de processamento, no entanto a probabilidade de filtrar informação útil sobre a captura é muito menor uma vez que já atua diretamente sobre a nuvem de pontos. Apesar de não conseguir eliminar todo o ruído, com este processamento é possível eliminar a maioria dos *outliers*, criando assim uma nuvem de pontos mais limpa. Apesar de ter bons resultados na análise e remoção de ruído de pequenas dimensões, esta técnica falha em zonas que tenham ruído de grandes dimensões pois estes são considerados informação útil. Como este processo requere uma análise extensiva da nuvem de pontos, a *performance* da aplicação desce aproximadamente 93% atingindo taxas de atualização 2/5 *frames* por segundo. A terceira solução pode ser considerada um último recurso pois permite escolher manualmente quais os pontos a eliminar e como tal, remover totalmente o ruído.

Já no terceiro caso, quando existe excesso de informação numa determinada zona, o ruído pode ser minimizado ou removido através da redução do número de pontos presentes na nuvem de pontos nessas áreas. A biblioteca *PCL* disponibiliza métodos para efetuar este filtro através da utilização de vóxeis, isto é, grelhas regulares tridimensionais que permitem dividir a nuvem

de pontos no espaço. Para cada vóxel são analisados os pontos nele contidos e substituídos por um ou mais pontos representativos, permitindo assim a diminuição do número de pontos e consequentemente, o ruído. O tamanho de cada vóxel, assim como o número de pontos representativos, são configuráveis, podendo assim este método ser adaptado a diferentes circunstancias, como o tamanho dos objetos ou a quantidade de ruído. No entanto, e assim como os outros métodos que atuam diretamente sobre a nuvem de pontos, este é um processo complexo e envolve um grande esforço computacional e como tal, não pode ser executado em tempo real. Os testes realizados permitiram verificar uma degradação da *performance* do sistema de cerca de 90% baixando a taxa de atualização para aproximadamente 5 *frames* por segundo.

4.3.3 Falhas de informação

Outra dificuldade, já referida na secção Configuração 360°, consiste na interferência entre as diferentes perspetivas de captura. Devido às reflexões causadas pelos pontos de aquisição virtuais, a emissão e sobreposição do padrão infravermelho da *Kinect* das várias perspetivas faz com que o valor de profundidade não seja calculado e, como tal, também não seja representado no mapa de profundidade. Nos testes realizados observou-se que este fenómeno ocorre essencialmente quando uma determinada superfície é capturada por mais que um ponto de captura (Figura 4.17). Como se pode observar nas zonas assinaladas com (1), nas perspetivas capturadas pelos espelhos, as áreas do chão encontram-se sem informação de profundidade resultado desta interferência. Este caso não é crítico uma vez que esta informação também seria descartada no entanto, noutras zonas como as assinaladas com (2) e (3), existe informação pertencente ao objeto que é perdida, influenciando o resultado final da captura. Este tipo de ruído não tem origem apenas nas interferências entre perspetivas. Caso os materiais dos objetos sejam reflexivos ou tenhas zonas transparentes, a *Kinect* pode também não conseguir capturar corretamente a informação de profundidade dessas áreas (zona (4)).

Para recuperar essa informação seria necessário reconstruir essas áreas. Uma vez que não é utilizada qualquer informação prévia sobre o objeto, este processo torna-se impossível de realizar de forma automática e/ou em tempo real. Foram testados métodos de processamento 2D simples, recorrendo a erosões de forma a estender a superfícies nessas áreas, contudo, como não se sabe nenhuma informação da geometria a adquirir, não é possível limitar a área de erosão. Isso faz com que também existam alterações nos limites dos objetos e, como tal, este método foi descartado. Durante os testes foi também observado que o ruído provocado pela interferência existe durante a captura mas, se houver movimento por parte do sensor, este atinge valores mínimos ou

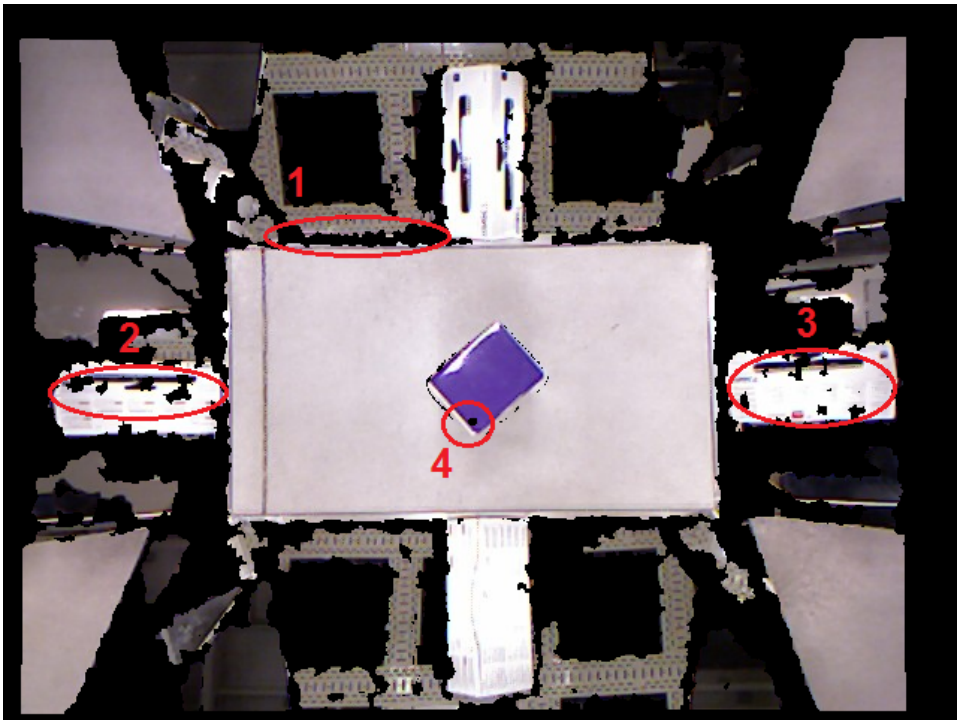


Figura 4.17: Demonstração dos diferentes tipos de Falhas de Informação durante a captura da Caixa da Kinect. *Em (1), interferência com chão, (2), informação perdida com reflexões, (4), materiais transparentes/reflexivos.*

desaparece totalmente. Esta técnica é usada de forma sistemática em [Schröder et al., 2011] para resolver este problema, no entanto não foi implementada no presente sistema uma vez que acarretava custos adicionais e poderia influenciar os detalhes nos resultados.

4.4 Tecnologia

Durante a descrição do sistema foram referidas e utilizadas várias tecnologias e bibliotecas para realizar diversas funções. Nesta secção pretende dar-se a conhecer um pouco melhor quais as ferramentas e *frameworks* usadas no desenvolvimento deste projeto, e qual a sua principal utilidade.

4.4.1 *OpenNI*

O *OpenNI* (*Open Natural Interaction*)[2] é uma biblioteca multiplataforma *Open Source* que serve de interface e permite aceder a vários dispositivos de aquisição

[2] `http://structure.io/openni` (acedido em Outubro de 2014)

de informação de profundidade, como é o caso da *Kinect* ou da *Asus XTion*. Esta biblioteca permite configurar diversos parâmetros dos sensores como, por exemplo, os de calibração entre imagem de profundidade e *RGB*. Neste projeto, o *OpenNI* foi utilizado para aceder à *Kinect* e retirar toda a informação disponível relativa à imagem de cor e ao mapa de profundidade. Além disso, e uma vez que este contém as propriedades de calibração do sensor, a conversão do mapa de profundidade em nuvem de pontos com as coordenadas do mundo real é também realizada a partir desta biblioteca.

4.4.2 OpenCV

O *OpenCV* (*Open Computer Vision*)[3] é uma biblioteca multiplataforma *Open Source* para processamento de imagem. Esta biblioteca está orientada para aplicações em tempo real que usam visão por computador e, como tal, foi desenvolvida a pensar no desempenho, oferecendo a possibilidade de tirar partido de processamento paralelo em múltiplos núcleos ou utilizando tecnologias mais específicas, como o *IPP* ou *CUDA*. Apesar de ter *wrappers* para outras linguagens, o *OpenCV* foi desenvolvido em *C/C++* e tem implementados vários métodos de processamento de imagem, cobrindo áreas como o reconhecimento de padrões, *machine learning*, etc. Neste projeto esta biblioteca é usada intensivamente desde o armazenamento da informação das imagens e mapa de profundidades da *Kinect*, até à sua visualização, passando por todo o processamento realizado em 2D, como a criação e utilização de máscaras e a aplicação de filtros de suavização.

4.4.3 PCL

O *Point Cloud Library* (PCL)[4] é um projeto multiplataforma *Open Source* para processamento de nuvens de pontos. Esta biblioteca, construída em *C++*, tem implementados vários algoritmos recentes para realizar esse processamento (como os utilizados para a remoção de ruído em 3D) e foi desenhada a pensar na *performance* e eficiência dos cálculos. Apesar de não ter sido utilizado, o *PCL* contém ainda *wrappers* para o *OpenNI* para realizar a aquisição de informação permitindo aceder a diferentes dispositivos inclusive de forma remota, ou seja, capturar informação de dispositivos presentes noutros computadores. Outra tecnologia presente nesta biblioteca é a do *VTK* (*Visualization Toolkit*), que está integrada num visualizador de informação, o *PCD Viewer*. Esta ferramenta específica foi utilizada neste projeto e permite a visualização e navegação em tempo real na nuvem de pontos em 3D. f

[3] http://opencv.org (acedido em Outubro de 2014)
[4] http://pointclouds.org/ (acedido em Outubro de 2014)

4.4.4 Outras

Além destas bibliotecas consideradas "essenciais" foram ainda utilizadas outras bibliotecas no desenvolvimento deste projeto. As mais relevantes foram:

- O *Boost*[5], uma biblioteca *C++ Open Source* com amplas funcionalidades e desenvolvida para melhorar a produtividade dos programadores. Foi utilizado para a criação do sistema *multi-threading* e para resolver os consequentes problemas de concorrência.

- O *Qt*[6], uma plataforma multiplataforma *Open Source* desenvolvida em *C++* que possui uma ferramenta de criação de interfaces gráficas, o *Qt Creator*. Foi utilizado para a construção da interface gráfica do sistema.

- O *MeshLab*[7], uma ferramenta *Open Source* que permite a visualização, o processamento e edição de nuvens de pontos e malhas poligonais 3D. Foi utilizado para visualização e edição de nuvem de pontos e ainda para análise e validação de resultados.

4.5 Sumario

De forma a concretizar os objetivos propostos foram construídos dois protótipos para realizar a captura 360°. O primeiro protótipo utiliza uma *Kinect* e dois espelhos, criando desta forma três pontos de captura distintos. Os elementos estão dispostos a distâncias semelhantes da área de captura formando um triângulo, centrado nessa área, entre eles. O segundo protótipo utiliza o mesmo sensor e quatro espelhos, criando como tal cinco pontos de captura. Nesta configuração os elementos estão dispostos entre si em forma de pirâmide com o centro da área de ação a coincidir com o centro da base da mesma.

A nível de fluxo de execução, o sistema é independente da topologia física utilizada e segue uma lógica de quatro passos. O primeiro passo é realizado uma única vez e é responsável pela configuração do sistema e os outros três, aquisição, processamento e visualização, constituem o ciclo de captura de informação:

- Configuração - definição da área de captura e das superfícies dos espelhos, assim como a sua calibração.

[5] www.boost.org (acedido em Outubro de 2014)
[6] http://qt-project.org (acedido em Outubro de 2014)
[7] http://meshlab.sourceforge.net (acedido em Outubro de 2014)

- Aquisição e pré-processamento - aquisição da informação de cor e profundidade da *Kinect*, processamento da mesma de forma a filtrar o máximo de informação desnecessária e geração da nuvem de pontos 3D.

- Processamento de informação - tratamento da imagem 2D e da nuvem de pontos de forma a melhorar o resultado de aquisição através da suavização de informação e remoção de *outliers*.

- Visualização de informação - visualização e navegação em tempo real da nuvem de pontos em 3D e análise dos resultados para validação dos mesmos.

Durante a implementação deste sistema foram detetados vários problemas relacionados com a aquisição de informação de profundidade e a qualidade da mesma. Os que mais se fizeram sentir foram a informação imprecisa (níveis de detalhe da informação de profundidade dada pela *Kinect*), ruído (informação do cenário mas não pertencente ao objeto), e falhas de informação (zonas do mapa de profundidade sem informação). Apesar de não ter sido possível resolver todos estes problemas, e ao mesmo tempo manter uma captura com taxas de atualização interativas, alguns deles foram abordados de forma a minimizar os seus efeitos sem perder outras propriedades.

Neste capítulo foram ainda descritas de forma breve quais as tecnologias utilizadas para o desenvolvimento deste projeto pelo que as mais importantes foram o *OpenNI, OpenCV e PCL*.

Capítulo 5

Resultados

Com o objetivo de validar o sistema desenvolvido foram selecionados alguns objetos para realizar a sua captura 360°. Dependendo das suas dimensões, essa aquisição foi efetuada por uma das duas configurações descritas na Secção 4.1. Além do desempenho genérico da aplicação em cada uma das situações, os resultados foram ainda avaliados para confirmar se estas correspondem às características dos objetos reais e para determinar a fiabilidade da captura.

Desta forma, e antes de entrar na análise de casos específicos, serão explicados quais os objetos a analisar, os parâmetros a avaliar e como é que essa avaliação foi realizada. Depois, o primeiro passo desta análise passa por validar os resultados, isto é, certificar que as nuvens de pontos capturadas respeitam a geometria original dos objetos reais. De seguida essa mesma nuvem será examinada de forma a aferir a qualidade da captura e a fiabilidade da mesma de acordo com o objeto original. A *performance* do sistema será analisada no final. Este capítulo será concluído com a análise dos resultados obtidos seguido de um resumo dos tópicos abordados.

5.1 Métodos de avaliação

De forma a conseguir aferir a qualidade do sistema desenvolvido foram selecionados alguns objetos e realizou-se a aquisição 360° dos mesmos. Esta avaliação consistiu na comparação entre os resultados obtidos com os objetos reais, tanto nas suas medidas como nas aproximações das *meshes* capturadas às superfícies originais. Para estes testes foram utilizados os seguintes objetos:

	Caixa da *Kinect*
	Dimensões: - Altura (38cm) - Largura (12cm) - Profundidade (15cm)
	Configuração Utilizada: - Caso 2
	Notas: Caixa com forma de um paralelepípedo de pequena dimensão. Constituída por cartão plastificado pelo que apresenta algum brilho, o que, em alguns momentos, dificulta a aquisição da informação de profundidade.

Figura 5.1: Informações do objeto "Caixa da Kinect *".*

	PC
	Dimensões: - Altura (43cm) - Largura (20cm) - Profundidade (50cm)
	Configuração Utilizada: - Caso 1
	Notas: Objeto com forma de um paralelepípedo com dimensão média. A superfície é lisa nas laterais e no topo no entanto a frente apresenta alguma curvatura.

Figura 5.2: Informações do objeto "PC".

	Bola
	Dimensões: - Diâmetro (24cm)
	Configuração Utilizada: - Caso 2
	Notas: Objeto com forma esférica com dimensão média. Toda a superfície é rugosa e opaca.

Figura 5.3: Informações do objeto "Bola".

Foram realizadas várias capturas destes objetos tendo cada uma delas características diferentes relativamente aos métodos utilizados durante a aquisição. A primeira característica prende-se com o tempo da captura, isto é, se esta foi feita de forma instantânea (uma *frame*) ou se foi uma captura prolongada que, para os testes propostos, foram consideradas 30 *frames*. As outras

características estão relacionadas com a utilização de filtros, tanto o filtro bilateral para a suavização das superfícies (descrito na Secção 4.3.2), como o filtro para a remoção de *outliers* em 3D (descrito na Secção 4.3.1). No caso da bola foi ainda feita uma segunda captura contemplando a movimentação da mesma. Neste caso apenas foi analisada a aplicação dos filtros, uma vez que um tempo de captura mais prolongado não se poderia aplicar aqui.

Em todos os casos o resultado das capturas foi guardado e exportado para um ficheiro *ply* para ser examinado posteriormente. Estes ficheiros assim como algumas imagens capturadas pela *Kinect*podem ser encontradas neste link[1]. A nomenclatura com três dígitos presentes no fim de cada ficherio correspondem, respetivamente, ao número de *frames* capturadas ('0' corresponde a uma *frame*, '1' corresponde a 30), a utilização do filtro de suavização Bilateral ('0' corresponde a desligado, '1' corresponde a ligado), e a utilização do filtro de remoção de *outliers* ('0' corresponde a desligado, '1' corresponde a ligado).

A análise dos resultados foi feita de forma a avaliar as seguintes características:

- Validação das dimensões dos modelos capturados e comparação com as do objeto original.
- Avaliação da qualidade dos resultados, isto é, saber o quão fiel e limpo o resultado é e quantificar ruídos e imperfeições verificadas.
- Desempenho da aplicação na geração de informação de acordo com os métodos e filtros utilizados.

Estas características serão descritas com mais detalhe nas respetivas secções. Para fazer-se esta análise utilizou-se o *MeshLab* uma vez que esta aplicação permite a visualização de nuvens de pontos e disponibiliza ferramentas que permitem fazer medições, seleções e alterações diretamente nas nuvens de pontos.

Os testes foram realizados utilizando um computador portátil com um processador *Intel®Core*™ i5-2410M @2.30GHz, 6Gb de memória RAM e com uma placa gráfica *NVIDIA GeForce GT 540M* com 1Gb de memória dedicados.

5.2 Validação de resultados

A validação dos resultados tem como objetivo comparar as dimensões dos objetos reais com a geometria obtida de forma a apurar a exatidão da captura.

[1] https://www.dropbox.com/sh/v8vjwvilbibchoq/AABdqm0RBpNWTaPyNEn5ZIZla?dl=0 (acedido em outubro de 2014)

Para tal foi considerada a geometria externa dos objetos o que, no caso dos objetos da Caixa da *Kinect* e do PC corresponde às suas arestas enquanto no caso da Bola, estática ou em movimento, corresponde ao seu diâmetro.

As medições efetuadas às capturas podem ser vistas nas tabelas seguintes. A linha "Real" corresponde à medição real efetuada sobre o objeto. As linhas "Mínimo" e "Máximo" representam as medições mínimas e máximas efetuadas sobre o modelo e são seguidas do valor relativo do desvio comparativamente com o valor real. A linha "Variação" corresponde à variação total entre as medições mínimas e máximas e à sua relação com as medidas reais.

Caixa da *Kinect*	Largura	Altura	Profundidade
Real	120mm	380mm	150mm
Mínimo	110mm	361mm	139mm
Desvio	10mm (8,33%)	19mm (5%)	11mm (7,33%)
Máximo	124mm	370mm	148mm
Desvio	4mm (3,33%)	10mm (2,63%)	2mm (1,33%)
Variação	14mm (11,67%)	29mm (7,63%)	13mm (8,67%)

Tabela 5.1: Validação de resultados, Caixa da Kinect

PC	Largura	Altura	Profundidade
Real	200mm	420mm	500mm
Mínimo	179mm	386mm	456mm
Desvio	21mm (10,5%)	34mm (8,1%)	44mm
Máximo	195mm	396mm	483mm (8,8%)
Desvio	5mm (2,5%)	24mm (5,71%)	17mm
Variação	26mm (13%)	58mm (13,8%)	61mm (12,2%)

Tabela 5.2: Validação de resultados, PC

Bola Estática	Diâmetro
Real	240mm
Mínimo	220mm
Desvio	20mm (8,33%)
Máximo	235mm
Desvio	5mm (2,08%)
Variação	25mm (10,42%)

Tabela 5.3: Validação de resultados, Bola Estática

Bola Movimento	Diâmetro
Real	240mm
Mínimo Desvio	215mm 25mm (10,42%)
Máximo Desvio	233mm 7mm (2,92%)
Variação	32mm (13,33%)

Tabela 5.4: Validação de Resultados, Bola em Movimento

Como se pode ver nos resultados, a diferença das dimensões entre os modelos capturados e os objetos reais é por norma pequena, atingindo os valores mais elevados durante a medição do PC. Neste caso o erro médio é de cerca de 13%. Um dos motivos para este erro é a utilização da configuração do Caso 1: o objeto encontra-se a uma distância maior da câmara e, como tal, o nível de detalhe é também menor induzindo assim medições menos precisas. De observar ainda que no caso da bola em movimento a diferença de dimensões também é mais elevada que a captura estática.

Outra conclusão que pode ser tirada destes resultados é que as medições apresentam um erro maior nas partes do objeto que são capturadas através dos espelhos. Este resultado já era esperado uma vez que a distância a que estas partes do objeto se encontram do sensor é maior e, como tal, estão mais suscetíveis a erros.

5.3 Avaliação de resultados

A avaliação da qualidade dos resultados consiste numa comparação entre os modelos capturados e os objetos reais, tanto a nível da forma como da consistência dos dados dos mesmos. Da fase de validação conclui-se que a geometria resultante das capturas, apesar de conter alguns erros, é fiel à dos objetos. Relativamente à consistência dos dados, foi necessário definir uma forma de quantificar essa característica e, para tal, olhou-se às dificuldades surgidas na fase de desenvolvimento e consideraram-se os problemas de informação imprecisa, ruído e falhas de informação.

5.3.1 Informação imprecisa

A imprecisão da informação está relacionada com a qualidade do próprio sensor e com as distâncias a que são efetuadas as capturas (descrito na Secção 4.3.1). Isto traduz-se na colocação dos pontos das superfícies dos objetos em posições

perto das corretas mas com desvios ligeiros.

Este fenómeno é mais evidente nas zonas com superfícies lisas. Como é esperado que a nuvem de pontos nestas áreas seja uniforme, qualquer perturbação visível é mais notória. Estas serão as zonas em foco para a avaliação da informação imprecisa.

Foram selecionadas algumas superfícies dos objetos com características diferentes, uma cuja aquisição tenha sido feita diretamente pelo sensor, e outra através dos espelhos, e nelas foi medida a distância à superfície dos pontos mais incorretos. As colunas "Medição" contêm o valor máximo medido, isto é, o valor mais impreciso em cada uma das áreas capturadas. As colunas "Melhoria" contêm o valor relativo da diferença na imprecisão da informação comparativamente com as medições sem filtro.

Frames Capturados	Filtro Suavização	*Outliers*	Direto		Espelho	
			Medição	Melhoria	Medição	Melhoria
1	Desligado	Desligado	13mm	-	76mm	-
1	Desligado	Ligado	12mm	7,69%	60mm	21,05%
1	Ligado	Desligado	9mm	30,77%	58mm	23,68%
1	Ligado	Ligado	8mm	38,46%	58mm	23,68%
30	Desligado	Desligado	10mm	-	76mm	-
30	Desligado	Ligado	7mm	30,00%	79mm	-3,95%
30	Ligado	Desligado	8mm	20,00%	73mm	3,95%
30	Ligado	Ligado	7mm	30,00%	65mm	14,47%

Tabela 5.5: Informação imprecisa, Caixa da Kinect.

Frames Capturados	Filtro Suavização	*Outliers*	Direto		Espelho	
			Medição	Melhoria	Medição	Melhoria
1	Desligado	Desligado	12mm	-	124mm	-
1	Desligado	Ligado	12mm	0,0%	105mm	15,32%
1	Ligado	Desligado	8mm	33,33%	101mm	18,55%
1	Ligado	Ligado	8mm	33,33%	99mm	20,16%
30	Desligado	Desligado	69mm	-	131mm	-
30	Desligado	Ligado	33mm	52,17%	122mm	6,87%
30	Ligado	Desligado	12mm	82,61%	120mm	8,4%
30	Ligado	Ligado	12mm	82,61%	120mm	8,4%

Tabela 5.6: Informação imprecisa, PC.

Como se pôde observar, a variação entre pontos de uma mesma superfície existe por toda ela e pode atingir os 131mm. Os casos mais graves registaram-se durante a captura do *PC* devido à maior distância a que este se encontra da câmara.

Frames Capturados	Filtro Suavização	*Outliers*	Direto		Espelho	
			Medição	Melhoria	Medição	Melhoria
1	Desligado	Desligado	21mm	-	52mm	-
1	Desligado	Ligado	11mm	47,62%	25mm	51,92%
1	Ligado	Desligado	12mm	42,86%	48mm	7,69%
1	Ligado	Ligado	11mm	47,62%	31mm	40,38%
30	Desligado	Desligado	25mm	-	61mm	-
30	Desligado	Ligado	13mm	8,0%	60mm	1,64%
30	Ligado	Desligado	19mm	24,0%	55mm	9,84%
30	Ligado	Ligado	19mm	24,0%	57mm	6,56%

Tabela 5.7: Informação imprecisa, Bola Estática.

Frames Capturados	Filtro Suavização	*Outliers*	Direto		Espelho	
			Medição	Melhoria	Medição	Melhoria
1	Desligado	Desligado	24mm	-	54mm	-
1	Desligado	Ligado	23mm	4,17%	31mm	42,59%
1	Ligado	Desligado	23mm	4,17%	34mm	37,04%
1	Ligado	Ligado	22mm	8,33%	33mm	38,89%

Tabela 5.8: Informação imprecisa, Bola em Movimento.

O modelo da Bola é o que tem menores variações devido à sua dimensão e ao material que a compõe. Como a superfície da bola é pouco reflexiva o padrão emitido pela *Kinect* é capturado mais fielmente o que se traduz em menos imprecisões. Comparativamente, o modelo da caixa da *Kinect* revela uma imprecisão maior que a bola uma vez que, apesar de se encontrarem à mesma distância, o material que a compõe é mais especular e como tal a aquisição de informação torna-se menos precisa.

Em qualquer um dos casos é notória a existência de uma imprecisão maior nas partes capturadas pelos espelhos uma vez que a distância a que estas se encontram do sensor é maior. A imprecisão da informação em relação à captura direta chega a ser 10 vezes superior.

A utilização de filtros revelou-se positiva na atenuação deste problema. Com a utilização do filtro de suavização a amplitude das imperfeições reduziram consideravelmente. Esta redução foi mais evidente nas zonas que eram capturadas diretamente pela câmara: uma vez que o erro era menor, a sua suavização permitiu melhorar a superfície capturada em média 36%. Nas zonas capturadas pelos espelhos, uma vez que existe mais ruído, a suavização não foi tão eficaz havendo uma melhoria média de 19%.

A utilização do filtro para a remoção de *outliers* também trouxe melhorias em relação à imprecisão da informação, no entanto numa grandeza menor. Nas zonas capturadas diretamente pela câmara a aplicação deste filtro trouxe

melhorias de cerca de 11%, enquanto que nas áreas capturadas através de espelhos essa melhoria foi de aproximadamente 15%. Estas últimas zonas contêm informação mais imprecisa e mais esparsa, como tal, a utilização do filtro de remoção de *outliers* é mais eficaz.

A aplicação dos dois filtros em simultâneo também produziu resultados positivos, atingindo assim uma maior melhoria na qualidade de informação. As áreas capturadas diretamente pela câmara registaram uma melhoria de cerca de 38% e as zonas capturadas através de espelhos 22%.

Por outro lado, a realização de uma captura prolongada revelou piores resultados a nível da imprecisão da informação. Embora visualmente as superfícies das nuvens de pontos fiquem mais suaves, a maior exposição leva a que seja também introduzida mais informação imprecisa esporádica. Este tipo de informação errada tem normalmente como origem as zonas de captura mais distantes, isto é, as zonas capturadas através dos espelhos. Isto faz com que existam zonas da *mesh* com maior variação posicional e, como tal, mais imprecisão.

5.3.2 Ruído

Nestas capturas o ruído representa excertos de informação que não foi possível filtrar corretamente e, como tal, fazem parte do resultado mas não deveriam fazer parte da nuvem de pontos do objeto. No capítulo anterior (Secção 4.3.2) foram identificados três tipos de ruído, no entanto no âmbito desta avaliação apenas teremos em conta os dois primeiros, isto é, as porções de informação que representam *outliers* e a informação incorreta.

Optou-se por descartar as zonas com excesso de informação uma vez que estas estão dependentes da forma do objeto e da configuração escolhida. Por exemplo, no caso da configuração do Caso 2, se se posicionar um objeto com uma geometria cúbica de forma às suas faces ficarem numa perspetiva paralela aos espelhos, não existirão zonas de sobreposição durante a aquisição e, como tal, também não existirão zonas com excesso de informação. Por outro lado, se o objeto não for colocado nessa posição, todas as faces laterais da caixa serão visíveis por mais que uma perspetiva e, como tal, todas elas terão zonas com excesso de informação.

A forma encontrada para avaliar esta característica passa pela quantificação do número de pontos que se enquadram nos erros descritos. A coluna "Total" contém o número total de pontos capturados nessa configuração enquanto a coluna "Ruído" contém os pontos dessa nuvem distinguidos como ruído. A coluna "Rácio" reflete a relação entre as duas colunas anteriores, isto é, a percentagem de informação ruidosa na captura.

Frames Capturados	Filtro Suavização	*Outliers*	Total	Ruído	Rácio
1	Desligado	Desligado	25122	1333	5,31%
1	Desligado	Ligado	24442	1117	4,57%
1	Ligado	Desligado	26272	1288	4,90%
1	Ligado	Ligado	25629	1118	4,36%
30	Desligado	Desligado	24683	2040	8,26%
30	Desligado	Ligado	24755	2004	8,10%
30	Ligado	Desligado	24821	1871	7,54%
30	Ligado	Ligado	24732	1986	8,03%

Tabela 5.9: Ruído, Caixa da Kinect.

Frames Capturados	Filtro Suavização	*Outliers*	Total	Ruído	Rácio
1	Desligado	Desligado	41672	1412	3,39%
1	Desligado	Ligado	40803	574	1,41%
1	Ligado	Desligado	41555	1342	3,23%
1	Ligado	Ligado	40846	626	1,53%
30	Desligado	Desligado	45900	5133	11,18%
30	Desligado	Ligado	45900	4826	10,51%
30	Ligado	Desligado	45722	5070	11,09%
30	Ligado	Ligado	45658	4659	10,20%

Tabela 5.10: Ruído, PC.

Frames Capturados	Filtro Suavização	*Outliers*	Total	Ruído	Rácio
1	Desligado	Desligado	16867	238	1,41%
1	Desligado	Ligado	16443	146	0,89%
1	Ligado	Desligado	16964	224	1,32%
1	Ligado	Ligado	16630	132	0,79%
30	Desligado	Desligado	19683	643	3,27%
30	Desligado	Ligado	19521	439	2,25%
30	Ligado	Desligado	19561	689	3,52%
30	Ligado	Ligado	19731	522	2,65%

Tabela 5.11: Ruído, Bola Estática.

Frames Capturados	Filtro Suavização	*Outliers*	Total	Ruído	Rácio
1	Desligado	Desligado	20076	866	4,31%
1	Desligado	Ligado	19506	564	2,89%
1	Ligado	Desligado	18848	774	4,11%
1	Ligado	Ligado	18860	624	3,31%

Tabela 5.12: Ruído, Bola em Movimento.

Analisando os resultados apresentados pode observar-se que a aplicação do filtro para remoção de *outliers* cumpre o seu propósito e como tal diminui a quantidade de ruído na maioria das capturas: até 7% em relação à captura sem filtro. A utilização do filtro de suavização tem também um efeito positivo em relação à quantidade de ruído, no entanto este efeito é reduzido atingindo apenas melhorias de 1%.

De notar ainda que a captura prolongada introduz mais ruído no modelo, principalmente na forma de *outliers*. Uma das características destes *outliers* é serem porções pequenas de informação não filtrada, muitas vezes resultado de erros esporádicos com origem no *hardware* na fase de captura. Numa aquisição instantânea (uma *frame*), este tipo de erros são mais improváveis de capturar e costumam ter como resultado fragmentos de informação de pequena dimensão. No entanto, no caso de uma captura prolongada, além de ser mais provável conseguir capturar este tipo de interferências, pode acontecer estas tornarem-se maiores por acumularem ruídos das várias *frames*. Estes casos dão origem a porções de informação com maior dimensão que, por sua vez, são mais difíceis de descartar automaticamente provocando um maior nível de ruído.

5.3.3 Falhas de informação

As falhas de informação resultam da inexistência de dados numa determinada área (Secção 4.3.3) o que provoca a ausência de informação nessas zonas na nuvem de pontos. Se considerarmos uma zona que seja capturada por duas perspetivas diferentes e apenas uma delas contiver falhas de informação, a geometria do objeto poderá ser complementada através da informação proveniente da outra perspetiva. No entanto, se as duas perspetivas contiverem falhas de informação, existirá na mesma uma área sem dados na nuvem de pontos e, uma vez que estas zonas estão expostas ao padrão da *Kinect* por mais que uma vez, é normal isto acontecer (Secção 3.2.2).

De forma a avaliar este ponto e tendo em conta as suas condicionantes, o método escolhido passou por uma análise do mapa de profundidades e pelo cálculo da razão entre o número total de pixéis da área do objeto e os pixéis sem

informação nessa mesma área. Essas máscaras foram construídas manualmente a partir da imagem de cor da captura como está representado na figura 5.4. No caso do objeto Bola em Movimento foram construídas duas máscaras, uma para cada uma das capturas.

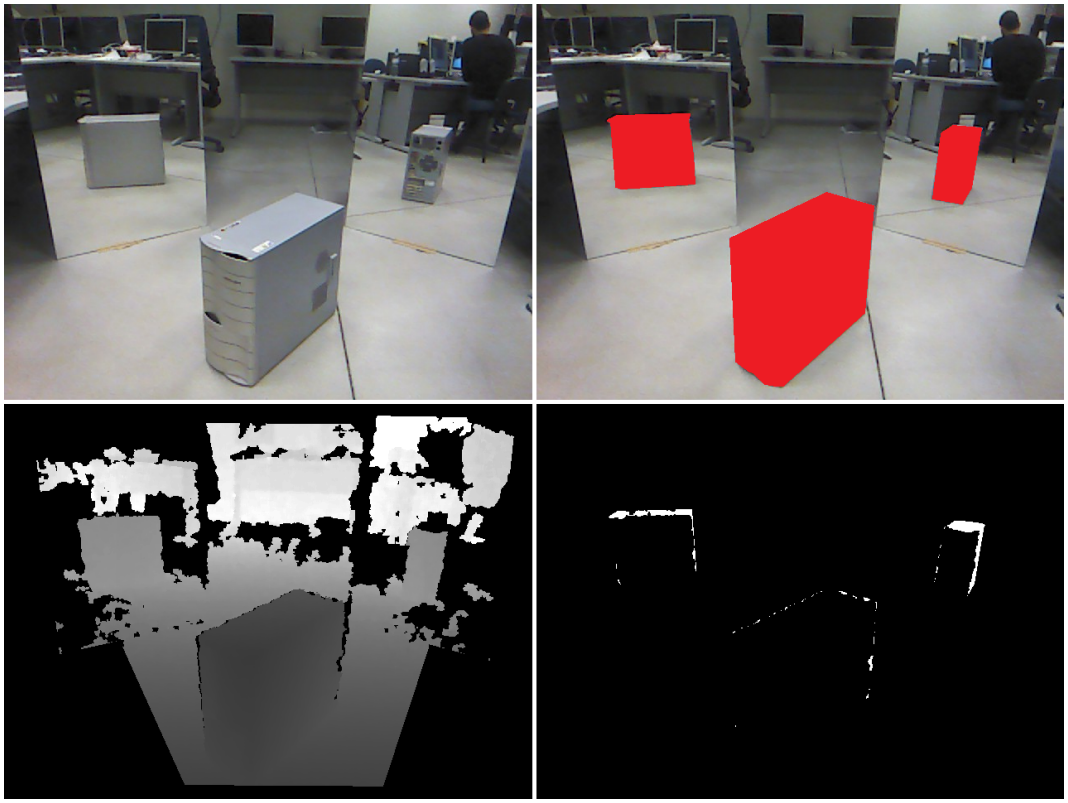


Figura 5.4: Imagens exemplificativas do processo para a obtenção do nível de Falhas de Informação. À esquerda, as imagens de cor (em cima) e de profundidade (em baixo) provenientes da captura. À direita, máscara construída a partir da imagem de cor (em cima) e a aplicação dessa mesma máscara à imagem de profundidade (em baixo) resultando nas zonas com falhas de informação.

Para este teste foram descartadas as capturas com o filtro de remoção de *outliers* uma vez que este apenas atua *a posteriori* sobre a nuvem de pontos 3D e não no mapa de profundidades. A coluna "Falhas" contém o número de pixéis da área do objeto sem informação e a coluna "Rácio" a relação entre estes e o número total de pixéis da máscara, ou seja, a percentagem de informação ruidosa contida na captura.

Máscara Total	*Frames* Capturados	Filtro Suavização	Ruído (píxeis)	Rácio
27751	1	Desligado	2397	8,64%
	1	Ligado	2444	8,81%
	30	Desligado	969	3,49%
	30	Ligado	1013	3,65%

Tabela 5.13: Falhas de informação, Caixa da Kinect.

Máscara Total	*Frames* Capturados	Filtro Suavização	Ruído (píxeis)	Rácio
43414	1	Desligado	1718	3,96%
	1	Ligado	1841	4,24%
	30	Desligado	838	1,93%
	30	Ligado	862	1,99%

Tabela 5.14: Falhas de informação, PC.

Máscara Total	*Frames* Capturados	Filtro Suavização	Ruído (píxeis)	Rácio
19376	1	Desligado	2784	14,37%
	1	Ligado	2741	14,15%
	30	Desligado	1043	5,38%
	30	Ligado	1030	5,32%

Tabela 5.15: Falhas de informação, Bola.

Máscara Total	*Frames* Capturados	Filtro Suavização	Ruído (píxeis)	Rácio
19718	1	Desligado	2162	10,96%
19611	1	Ligado	2098	10,70%

Tabela 5.16: Falhas de informação, Bola em Movimento.

Como se pode observar nos resultados obtidos, a área com falhas de informação é menor nas capturas prolongadas comparativamente com as capturas instantâneas, podendo esta redução chegar a aproximadamente 9%. As falhas de informação são muitas vezes inconsistentes, isto é, não aparecem em todas as *frames*. Assim, numa captura prolongada a nuvem de pontos é construída a partir da acumulação de informação de várias *frames* e como tal, as falhas esporádicas que ocorrem numa *frame* são corrigidas nas *frames* seguintes. Já o filtro de suavização tem um efeito quase nulo, pelo que a sua aplicação apenas tem alterações no nível do ruído na ordem dos 0,17%.

5.4 Desempenho

O desempenho da aplicação está intimamente ligado à quantidade de informação a processar e ao processamento efetuado sobre essa informação: quanto maior a quantidade de informação e quantos mais filtros forem usados, mais processamento é necessário para calcular os resultados e consequentemente, menor o desempenho do sistema.

Os filtros utilizados durante os testes foram o filtro bilateral para a suavização das superfícies do objeto e o filtro de remoção de *outliers*. Na tabela seguinte são apresentadas as taxas de atualização (*frames* por segundo) obtidas com e sem a aplicação de filtros. É também mostrada a penalização relativa do desempenho do sistema com a utilização dos filtros.

	Sem Filtros	Filtro Bilateral	Filtro *Outliers*	Ambos
PC (43507pts)	45,10 fps	26,21 (41,87%)	2,51 (94,44%)	2,36 (94,78%)
Kinect (25057pts)	49,62 fps	28,42 (42,73%)	3,52 (92,92%)	3,37 (93,22%)
Bola (18175pts)	62,94 fps	31,56 (49,85%)	5,42 (91,38%)	5,01 (92,04%)

*Tabela 5.17: Desempenho registado pelas diferentes configurações (*frames *por segundo)*

Como já tinha sido referido na descrição de cada um dos filtros, o desempenho do sistema é penalizado em cerca de 45% com a utilização do filtro Bilateral, 93% com a utilização do filtro de Remoção de *Outliers* e 94% se forem utilizados os dois filtros em simultâneo. Nos casos em que é utilizado o filtro de Remoção de *Outliers*, uma vez que exige mais processamento, o desempenho do sistema deixa de conseguir atingir taxas de atualização interativas e como tal, este filtro apenas pode ser usado para a captura de objetos estáticos.

5.5 Análise de resultados

Durante os testes efetuados foi analisada a qualidade dos resultados em relação à fiabilidade das dimensões capturadas e das superfícies.

Relativamente à validação das dimensões, o erro encontrado não foi muito elevado, atingiu no máximo os 14%, no entanto é o suficiente para não ser possível uma reprodução fiel dos objetos originais. Isto representa um erro de quase 6cm. Estas irregularidades encontraram-se principalmente nas zonas cuja superfície do objeto se encontra mais longe do sensor uma vez que o erro associado à captura também aumenta. Isto provoca um maior desnivelamento e mais falhas na informação, o que leva a medições menos corretas.

Em relação à qualidade da nuvem de pontos, esta também ficou um pouco aquém do esperado. Nas perspetivas capturadas diretamente pela *Kinect*, uma vez que a distância das superfícies ao sensor é menor, o erro é relativamente pequeno e o resultado satisfatório. O desvio maior ocorreu na captura de um objeto em movimento e registou um erro de 2,4cm. No entanto, nas zonas capturadas a partir dos espelhos, a distância de captura é maior e, acompanhando a fórmula de erro da *Kinect* (descrita na Secção 3.3.2), a imprecisão da informação é bastante elevada chegando a atingir no pior caso desvios de 13cm. Este registo aconteceu durante a captura do PC onde as superfícies capturadas pelos espelhos encontravam-se a distâncias superiores a 3m. Estes resultados revelaram-se insatisfatórios para a construção de modelos 3D de objetos com boa qualidade.

Quanto ao desempenho, a aplicação obteve bons resultados. Apesar do acréscimo de processamento para computar os resultados provenientes dos espelhos, sem a aplicação de filtros o sistema apresentou taxas de atualização superiores a 45 *frames* por segundo. Neste caso a principal limitação é a velocidade de captura da *Kinect*. No entanto, com a aplicação dos filtros o desempenho do sistema baixou consideravelmente: cerca de 40% com a utilização do filtro de suavização bilateral e mais de 90% com a utilização do filtro de remoção de *outliers*. Isto significa que, no *hardware* utilizado, as taxas de atualização baixaram para menos de 5 *frames* por segundo impossibilitando assim o funcionamento do sistema em tempo real.

Assim como foi mostrado, o principal ponto de influência nos resultados de captura foi a distância a que as superfícies dos objetos se encontravam da fonte de captura. Devido ao erro exponencial relatado na Secção 3.3.2, superfícies que se encontram a 150cm de distância já contêm uma variação de 1,7cm o que é bastante visível tendo em conta a dimensão dos objetos. Como todas as superfícies capturadas pela perspetiva do espelho se encontram a mais de 250cm, a nuvem de pontos nessas zonas tem pouca qualidade.

Contudo e dado o aspeto geral dos modelos, pode observar-se que a forma mantém-se e são percetíveis os traços gerais do objeto. Embora os erros encontrados impossibilitem uma reprodução fiel do objeto capturado, é possível construir aproximações desses mesmos objetos. Se a isso aliarmos a velocidade do sistema sem a aplicação de filtros, a aplicação da informação gerada a sistemas interativos pode possibilitar a construção de esqueletos a partir dessas aproximações.

Foram feitos também alguns testes de forma a avaliar o desempenho do sistema na geração em tempo real das malhas poligonais. Para tal foram utilizados os algoritmos de triangulação de superfície gananciosos (*Greedy Surface Triangulation*)[Marton et al., 2009] disponibilizados pela biblioteca *PCL*, no entanto, o resultado não foi satisfatório (Figura 5.5).

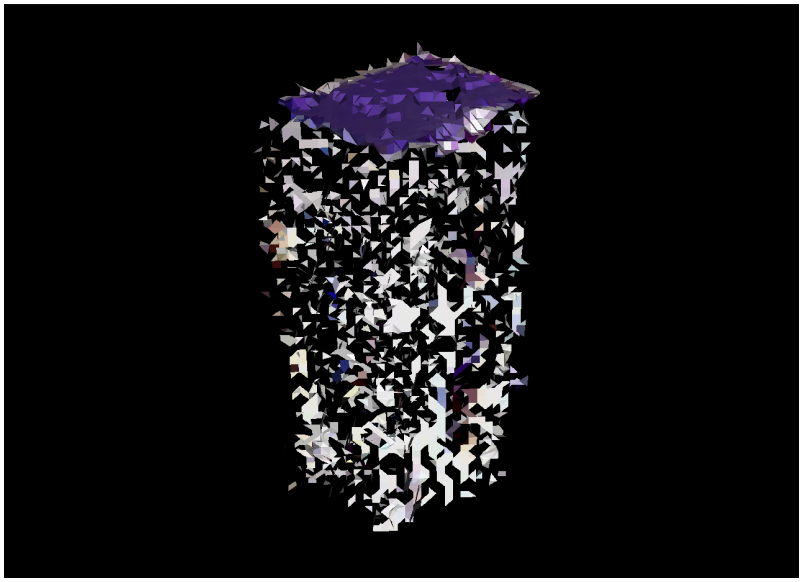


Figura 5.5: Exemplo dos resultados obtidos na tentativa de geração de uma malha poligonal a partir da nuvem de pontos capturada.

Este algoritmo utiliza uma técnica de vizinho mais próximo, isto é, para cada ponto é calculada a vizinhança e todos os pontos contidos nessa seleção são analisados para inclusão na *mesh*. É possível variar alguns parâmetros que influenciam os resultados do algoritmo. No entanto, uma vez que a quantidade de pontos que constituem a superfície de um objeto é variável e inconstante, não foi possível encontrar os valores ótimos para a correta geração da malha poligonal em todas as superfícies do objeto. Além disso, a *performance* do sistema baixava para taxas de atualização de uma *frame* por segundo o que anularia por completo ao sistema a propriedade de aquisição em tempo real.

5.6 Sumário

De forma a avaliar o sistema construído foram selecionados alguns objetos de teste: um PC, a Caixa da *Kinect* e uma Bola de Basquete. O primeiro objeto foi capturado usando a configuração do Caso 1 e para os outros dois recorreu-se à configuração arena. Depois de capturada a nuvem de pontos a análise dos resultados passou pela validação das suas dimensões, avaliação da qualidade das superfícies capturadas e pela análise do desempenho do sistema.

Relativamente às dimensões, os erros observados, apesar de pequenos, atingiram imprecisões de cerca de 14% em relação aos objetos originais. Apesar da forma se manter e ser percetível a geometria do objeto, um erro desta ordem impossibilita a reprodução fiel dos objetos capturados. Quanto à qualidade das

superfícies capturadas, as imprecisões registadas foram também elas elevadas chegando a atingir os 13cm. A aplicação de filtros revelou-se eficaz chegando em alguns casos a apresentar melhorias superiores a 30%. Contudo isso não foi o suficiente para criar superfícies fiéis às dos objetos originais. Todas elas se apresentaram bastante ruidosas. Em qualquer um dos casos, a distância a que a superfície do objeto se encontra do sensor revelou-se decisiva na qualidade dos resultados. As superfícies que foram capturadas diretamente pelo sensor apresentam uma qualidade superior às capturadas através dos espelhos uma vez que se encontram mais próximas do mesmo.

A nível de desempenho o sistema apresentou bons resultados atingindo taxas de atualização superiores a 45 *frames* por segundo. Estes valores baixam com a aplicação dos filtros usados para o aperfeiçoamento da nuvem de pontos, cerca de 40% no caso do filtro de suavização Bilateral e chegando aos 90% com o filtro de Remoção de *Outliers*.

Apesar da qualidade dos resultados ter ficado um pouco aquém do esperado, a forma e dimensões gerais do objeto mantêm-se. Se aliarmos isso ao desempenho atingido é possível a criação de modelos 3D com aproximações dos objetos reais com taxas de atualização interativas. Isto pode ser usado de diversas formas em sistemas reativos e interativos nomeadamente a partir da geração do esqueleto 3D da entidade capturada.

Capítulo 6

Conclusão e trabalho futuro

Esta tese teve como objetivo o estudo e implementação de um sistema de baixo custo, capaz de realizar a captura 360° de informação 3D sobre um objeto em tempo real, utilizando uma configuração estática. Este sistema seria capaz de capturar toda a informação geométrica sobre a superfície de um objeto ao longo de todo o seu perímetro exterior, isto é, conforme vista de qualquer ponto de um círculo que contenha o objeto no seu interior.

As características descritas nem sempre são compatíveis, como tal, foi necessário criar uma solução capaz de as integrar todas num sistema único. Foram estudadas várias abordagens, tanto a nível do sensor de captura, como da configuração utilizada, e a escolha recaiu sobre a *Microsoft Kinect* e uma configuração de espelhos. A *Kinect* tem como principal vantagem o acesso a informação 3D sobre o mundo em tempo real a um custo reduzido. A configuração de espelhos permite que informação que não está no raio de visão do sensor fosse capturada, o que possibilita a captura 360°. Esta configuração é estática, no entanto, versátil uma vez que pode ser adaptada consoante a dimensão dos objetos. A principal limitação deste sistema recai sobre a distância a que a *Kinect* consegue recolher informação 3D, o que restringe a colocação dos espelhos e, consequentemente, a dimensão dos objetos capturados.

Para testar o sistema desenvolvido foram escolhidos três objetos e recorreu-se a duas configurações distintas, como foi descrito no capítulo Arquitetura do sistema: uma utilizando a *Kinect* e dois espelhos, criando desta forma três pontos de captura distintos, e outra tirando partido do mesmo sensor e quatro espelhos, criando assim cinco pontos de captura.

A nuvem de pontos foi capturada para cada um dos objetos e analisada de forma a apurar a qualidade dessa aquisição. Para tal foram avaliados três pontos: validação da dimensão do objeto, qualidade da malha capturada e desempenho do sistema. Os resultados obtidos não foram tão bons como o esperado. A validação da dimensão do objeto apresentou erros na ordem dos

14% que, apesar de não ser muito elevado, impossibilita a recriação fiel dos objetos originais. Quanto à qualidade das superfícies capturadas, registou-se bastante ruído e imprecisão na informação chegando a atingir variações de 13cm. Apesar destes erros terem sido minorados com a aplicação de filtros de suavização e de remoção de *outliers*, esta melhoria não foi suficiente e os resultados continuaram com muita imprecisão. Além disso, a aplicação destes filtros penalizou bastante o desempenho do sistema: 40% no caso do filtro de suavização e mais de 90% no filtro de remoção de *outliers*. Neste campo o desempenho do sistema foi bom e, sem a utilização desses filtros, atingiu taxas de atualização superiores a 45 *frames* por segundo. Estes valores são afetados pela dimensão do objeto pelo que no caso dos objetos mais pequenos chegou a atingir-se taxas de 60 *frames* por segundo.

Dos resultados obtidos, o principal problema e limitação do sistema prendeu-se com a distância a que os objetos se encontram do sensor. Uma vez que o erro nas medições da *Kinect* está associado a uma função quadrática (ver Secção 3.3.2), com o aumentar da distância, o erro também aumenta: se o objeto se encontrar a 2,5m do sensor, o erro associado a essa medição já é de 4,8cm. Esta limitação teve consequências principalmente na informação capturada a partir dos espelhos. Uma vez que, nestes casos, a superfície do objeto se encontra mais distante do sensor (corresponde à soma da distância da superfície ao espelho e do espelho ao sensor), os resultados obtidos por esta via apresentaram pior qualidade.

Os objetos de maior dimensão obrigaram à construção de uma configuração mais ampla e, como tal, o sensor ficou mais afastado do objeto e dos espelhos. Consequentemente, a captura nestes casos apresentou piores resultados. No caso de objetos mais pequenos é possível aproximar todos os intervenientes e realizar a captura 360°. A área de ação tornou-se mais pequena e, como tal, a qualidade da nuvem de pontos capturada foi maior.

Com estes resultados conclui-se que o sistema não era apto para criar modelos de objetos 3D fieis aos objetos originais devido aos níveis de ruído e imprecisão registados. No entanto, as nuvens de pontos geradas respeitam a forma dos objetos e a sua dimensão também se aproxima da original. Aliando isso ao desempenho que a aplicação apresentou, este sistema pode ser usado para gerar aproximações 3D dos objetos em tempo real, o que pode ser útil para aplicações interativas ou outras aplicações que não precisem de um nível de detalhe elevado. A informação 360° em 3D sobre o corpo humano pode ser útil para sistemas *Virtual Try-On*, tanto de corpo inteiro como parcial, ou na criação de esboços de movimentos e recriação do esqueleto humano.

6.1 Trabalho futuro

Apesar dos resultados não terem sido os esperados foi possível demonstrar com sucesso este conceito. Existem vários pontos onde este sistema pode sofrer melhorias, tanto a nível de *hardware* como de processamento.

A menor qualidade das malhas capturadas provém maioritariamente do erro associado à captura por parte da *Kinect*. A utilização de um sensor que ofereça uma melhor qualidade da informação 3D a distâncias maiores melhoraria por si só os resultados do sistema. A utilização da *Kinect2* poderia ser uma solução a seguir uma vez que, além de conseguir gerar informação mais precisa e consistente, vem equipada com uma câmara *RGB FullHD*. Outra opção relacionada apenas com a captura de informação a curta distância poderia passar pela utilização do sensor *Leap Motion*: este sensor afirma ter precisões de movimento na ordem dos centésimos de milímetro. Apesar de não fazer a captura de informação *RGB*, se este for integrado e calibrado com um sensor externo poderia ser uma solução de alta precisão a um custo reduzido.

Outro ponto de falha no sistema construído é o processo de calibração. A obtenção do plano representativo da posição de cada espelho revelou-se difícil e a introdução de erro neste processo frequente. A utilização de métodos mais precisos, tanto através da medição das distâncias entre componentes como na utilização de uma área maior para representar o espelho, poderia facilitar o processo de calibração e também produzir resultados com mais qualidade. Com a melhoria destes dois pontos seria possível pensar na utilização das reflexões entre espelhos para gerar mais pontos de captura em perspetivas diferentes utilizando os mesmos recursos.

Relativamente ao desempenho do sistema, apesar de este ter sido satisfatório, poderia ainda ser otimizado através da utilização de *GPU*. Isto permitiria a utilização de mais filtros e filtros mais eficazes sem que o desempenho do sistema fosse tão penalizado. Este é um campo em foco no desenvolvimento das bibliotecas utilizadas neste projeto (*OpenCV* e *PCL*) e estes progressos estão a ser disponibilizados com frequência.

Num campo diferente e menos relacionado com o projeto, além da passagem de modelos estáticos para ficheiro, outra funcionalidade interessante poderia passar pela gravação de vídeo da informação já processada em 3D. O *OpenNI* permite a gravação do fluxo de informação da *Kinect* para ficheiros, mas este apenas contém a informação crua. Todo o processamento terá de ser efetuado sempre que se quiser reproduzir o vídeo. Isto torna este formato ineficiente e impossibilita que seja guardada informação já processada sobre, por exemplo, a malha poligonal dos objetos capturados. A gravação de informação 3D é um processo exigente a nível de espaço, uma vez que, além da imagem, também estão associados a malha poligonal, normais dos pontos, texturas, etc. Estão

a ser feitos esforços neste campo por parte da *PCL*[1] mas para já ainda não foram disponibilizados. A possibilidade de guardar a informação já filtrada, mesmo que não fosse em tempo real, seria uma mais-valia para o projeto.

Para terminar, o sistema desenvolvido necessita de melhorias em vários pontos para conseguir produzir resultados com níveis de qualidade satisfatórios. Além disso, também seria necessário mais trabalho a nível de consistência para o tornar usável para o público em geral. No entanto, o sistema está funcional e os progressos feitos foram positivos podendo dar origem a novas iterações.

[1] http://pointclouds.org/gsoc/ (acedido em outubro de 2014)

Referências

D.S. Alexiadis, D. Zarpalas, and P. Daras. Real-time, realistic full-body 3d reconstruction and texture mapping from multiple kinects. In *IVMSP Workshop, 2013 IEEE 11th*, pages 1–4, June 2013. doi: 10.1109/IVMSPW.2013.6611939. **Cited** on pages 22, 37 and 38.

Abdennour Aouina, Michel Devy, and Antonio Marín-Hernández. Comparison of active sensors for 3d modeling of indoor environments. In Jean-Louis Ferrier, Oleg Yu. Gusikhin, Kurosh Madani, and Jurek Z. Sasiadek, editors, *ICINCO (1)*, pages 442–449. SciTePress, 2013. ISBN 978-989-8565-70-9. URL http://dblp.uni-trier.de/db/conf/icinco/icinco2013-1.html#AouinaDM13. **Cited** on page 42.

F. Bellocchio and S. Ferrari. *Depth Map and 3D Imaging Applications: Algorithms and Technologies*, chapter 3D Scanner, State of the Art, pages 451–470. IGI Global, 2011. **Cited** on pages 7, 8, 9 and 11.

E. Berndt and J. Carlos. Cultural heritage in the mature era of computer graphics. *Computer Graphics and Applications, IEEE*, 20(1):36–37, Jan 2000. ISSN 0272-1716. doi: 10.1109/38.814549. **Cited** on page 5.

J. Boehm. Natural user interface sensors for human body measurement. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXIX-B3:531–536, 2012. doi: 10.5194/isprsarchives-XXXIX-B3-531-2012. URL http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XXXIX-B3/531/2012/. **Cited** on page 22.

G. Borenstein. *Making Things See: 3D Vision with Kinect, Processing, Arduino, and MakerBot*. Make: Books. O'Reilly Media, Incorporated, 2012. ISBN 9781449307073. URL http://books.google.de/books?id=G8Ua-YTU4CoC. **Cited** on page 38.

K.L. Boyer and A.C. Kak. Color-encoded structured light for rapid active ranging. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-9(1): 14–28, Jan 1987. ISSN 0162-8828. doi: 10.1109/TPAMI.1987.4767869. **Cited** on page 14.

L. Cruz, D. Lucio, and L. Velho. Kinect and rgbd images: Challenges and applications. In *Graphics, Patterns and Images Tutorials (SIBGRAPI-T), 2012 25th SIBGRAPI Conference on*, pages 36–49, Aug 2012. doi: 10.1109/SIBGRAPI-T.2012.13. **Cited** on page 17.

Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981. **Cited** on page 59.

Jason Geng. Structured-light 3d surface imaging: aătutorial. *Adv. Opt. Photon.*, 3 (2):128–160, Jun 2011. doi: 10.1364/AOP.3.000128. URL http://aop.osa.org/abstract.cfm?URI=aop-3-2-128. **Cited** on page 14.

Z. Jason Geng. Rainbow three-dimensional camera: new concept of high-speed three-dimensional vision systems. *Optical Engineering*, 35(2):376–383, 1996. doi: 10.1117/1.601023. URL http://dx.doi.org/10.1117/1.601023. **Cited** on page 14.

Peter Henry, Michael Krainin, Evan Herbst, Xiaofeng Ren, and Dieter Fox. Rgb-d mapping: Using depth cameras for dense 3d modeling of indoor environments. In *In the 12th International Symposium on Experimental Robotics (ISER*, 2010. **Cited** on pages 5 and 30.

Peter R.M. Jones and Marc Rioux. Three-dimensional surface anthropometry: Applications to the human body. *Optics and Lasers in Engineering*, 28(2):89 – 117, 1997. ISSN 0143-8166. doi: http://dx.doi.org/10.1016/S0143-8166(97)00006-7. URL http://www.sciencedirect.com/science/article/pii/S0143816697000067. Applications of the Automated Measurement of Human Size and Shape. **Cited** on page 5.

Bernhard Kainz, Stefan Hauswiesner, Gerhard Reitmayr, Markus Steinberger, Raphael Grasset, Lukas Gruber, Eduardo Veas, Denis Kalkofen, Hartmut Seichter, and Dieter Schmalstieg. Omnikinect: Real-time dense volumetric data acquisition and applications. In *Proceedings of the 18th ACM Symposium on Virtual Reality Software and Technology*, VRST '12, pages 25–32, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1469-5. doi: 10.1145/2407336.2407342. URL http://doi.acm.org/10.1145/2407336.2407342. **Cited** on page 23.

Dong-Ik Ko and Gaurav Agarwal. Gesture recognition: Enabling natural interactions with electronics, 2012. URL http://www-cs.ccny.cuny.edu/~wolberg/capstone/kinect/GestureRecognitionTI.pdf. **Cited** on page 8.

D. Lanman, D. Crispell, and Gabriel Taubin. Surround structured lighting for full object scanning. In *3-D Digital Imaging and Modeling, 2007. 3DIM '07. Sixth International Conference on*, pages 107–116, Aug 2007. doi: 10.1109/3DIM.2007.57. **Cited** on pages 22, 38 and 51.

T. P. Lerch, M. MacGillivray, and T. Domina. 3d laser scanning: A model of multidisciplinary research. *Journal of Textile, Apparel, Technology and Management*, 5, 2006. **Cited** on page 5.

Marc Levoy, Kari Pulli, Brian Curless, Szymon Rusinkiewicz, David Koller, Lucas Pereira, Matt Ginzton, Sean Anderson, James Davis, Jeremy Ginsberg, Jonathan Shade, and Duane Fulk. The Digital Michelangelo Project: 3D scanning of large statues. In *Proceedings of ACM SIGGRAPH 2000*, pages 131–144, July 2000. **Cited** on page 27.

J. Rurainsky M. Marcon. 3d face reconstruction from a single camera using a multi mirror set-up. *IET Conference Proceedings*, pages 8–8(1), January 2009. URL http://digital-library.theiet.org/content/conferences/10.1049/ic.2009.0236. **Cited** on page 22.

Zoltan Csaba Marton, Radu Bogdan Rusu, and Michael Beetz. On Fast Surface Reconstruction Methods for Large and Noisy Datasets. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Kobe, Japan, May 12-17 2009. **Cited** on pages 64 and 90.

Sven Molkenstruck, Simon Winkelbach, and FriedrichM. Wahl. 3d body scanning in a mirror cabinet. In Gerhard Rigoll, editor, *Pattern Recognition*, volume 5096 of *Lecture Notes in Computer Science*, pages 284–293. Springer Berlin Heidelberg, 2008. ISBN 978-3-540-69320-8. doi: 10.1007/978-3-540-69321-5_29. URL http://dx.doi.org/10.1007/978-3-540-69321-5_29. **Cited** on pages 25, 26 and 38.

Carlo Dal Mutto, Pietro Zanuttigh, and Guido M. Cortelazzo. *Time-of-Flight Cameras and Microsoft Kinect(TM)*. Springer Publishing Company, Incorporated, 2012. ISBN 1461438063, 9781461438069. **Cited** on pages 11, 12 and 17.

Richard A Newcombe, Andrew J Davison, Shahram Izadi, Pushmeet Kohli, Otmar Hilliges, Jamie Shotton, David Molyneaux, Steve Hodges, David Kim, and Andrew Fitzgibbon. Kinectfusion: Real-time dense surface mapping and tracking. In *Mixed and augmented reality (ISMAR), 2011 10th IEEE international symposium on*, pages 127–136. IEEE, 2011. **Cited** on page 23.

Pierre Payeur and Danick Desjardins. Structured light stereoscopic imaging with dynamic pseudo-random patterns. In Mohamed Kamel and Aurélio Campilho, editors, *Image Analysis and Recognition*, volume 5627 of *Lecture Notes in Computer Science*, pages 687–696. Springer Berlin Heidelberg, 2009. ISBN 978-3-642-02610-2. doi: 10.1007/978-3-642-02611-9_68. URL http://dx.doi.org/10.1007/978-3-642-02611-9_68. **Cited** on page 14.

ROS.org. Openni tutorials. http://wiki.ros.org/openni_launch/Tutorials, 2012. Último acesso em outubro de 2014. **Cited** on page 60.

Yannic Schröder, Alexander Scholz, Kai Berger, Kai Ruhl, Stefan Guthe, and Marcus Magnor. Multiple kinect studies. Technical Report 09-15, ICG, October 2011. **Cited** on page 72.

Jan Smisek, Michal Jancosek, and Tomás Pajdla. 3d with kinect. In *ICCV Workshops*, pages 1154–1160. IEEE, 2011. ISBN 978-1-4673-0062-9. URL http://dblp.uni-trier.de/db/conf/iccvw/iccvw2011.html#SmisekJP11. **Cited** on page 42.

Sebastian Thrun and John J Leonard. Simultaneous localization and mapping. *Springer handbook of robotics*, pages 871–889, 2008. **Cited** on pages 23 and 30.

C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proceedings of the Sixth International Conference on Computer Vision*, ICCV '98, pages 839–, Washington, DC, USA, 1998. IEEE Computer Society. ISBN 81-7319-221-9. URL http://dl.acm.org/citation.cfm?id=938978.939190. **Cited** on page 67.

Wikipedia. Espelho. http://pt.wikipedia.org/wiki/Espelho, 2014. Último acesso em outubro de 2014. **Cited** on pages 40 and 41.